Centre for International
Governance Innovation

CIGI Papers No. 187 — September 2018

Data Ownership

Teresa Scassa

Centre for International
Governance Innovation

CIGI Papers No. 187 — September 2018

Data Ownership

Teresa Scassa

CIGI Masthead

Printed in Canada on paper containing 100% post-consumer fibre and certified by the Forest Stewardship Council® and the Sustainable Forestry Initiative.

Centre for International Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

Table of Contents

About the Author

Teresa Scassa is a senior fellow with CIGI's International Law Research Program. She is also the Canada Research Chair in Information Law and Policy and a full professor at the University of Ottawa's Law Faculty, where her groundbreaking research explores issues of data ownership and control. Teresa is an award-winning scholar and is the author and editor of five books and more than 65 peer-reviewed articles and book chapters. She has a track record of interdisciplinary collaboration to solve complex problems of law and data and was part of the Geothink research partnership. Teresa 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.

At CIGI, Teresa's research focuses on the legal challenges associated with data ownership and the need for a national data strategy in a data-driven economy. Her research also covers the governance of smart cities' data and its implications for innovation, transparency, accountability, sovereignty and privacy.

Teresa has worked as a consultant for government and the private sector and has also worked with non-governmental organizations on issues within her areas of legal expertise. She is a past member of the External Advisory Committee of the Office of the Privacy Commissioner of Canada, and of the Canadian Government Advisory Committee on Open Government. Teresa holds degrees in civil and common law from McGill University, as well as an LL.M. and a doctorate from the University of Michigan. She clerked for Mme Justice Claire L'Heureux-Dubé at the Supreme Court of Canada from 1988 to 1989.

About the Program

The International Law Research Program (ILRP) at CIGI is an integrated multidisciplinary research program that provides leading academics, government and private sector legal experts, as well as students from Canada and abroad, with the opportunity to contribute to advancements in international law.

The ILRP strives to be the world's leading international law research program, with recognized impact on how international law is brought to bear on significant global issues. The program's mission is to connect knowledge, policy and practice to build the international law framework — the globalized rule of law — to support international governance of the future. Its founding belief is that better international governance, including a strengthened international law framework, can improve the lives of people everywhere, increase prosperity, ensure global sustainability, address inequality, safeguard human rights and promote a more secure world.

The ILRP focuses on the areas of international law that are most important to global innovation, prosperity and sustainability: international economic law, international intellectual property law and international environmental law. In its research, the ILRP is attentive to the emerging interactions among international and transnational law, Indigenous law and constitutional law.

Acronyms and Abbreviations

AI	artificial intelligence
BCFNDGI	British Columbia First Nations' Data Governance Initiative
GDPR	General Data Protection Regulation
IP	intellectual property
OCAP	ownership, control, access and possession
SCC	Supreme Court of Canada
TPMs	technological protection measures
TRIPS	Agreement on Trade-Related Aspects of Intellectual Property Rights
WCT	WIPO Copyright Treaty
WIPO	World Intellectual Property Organization
WPPT	WIPO Performances and Phonograms Treaty

Executive Summary

The rapid expansion of the data economy raises serious questions about who "owns" data and what data "ownership" entails. In most jurisdictions, data that are kept confidential can be protected as confidential information. However, such data are vulnerable to exposure through hacking or leaking by third parties. In many instances, significant stores of data cannot be kept confidential, and protection must be sought elsewhere. Copyright law has long treated facts as being in the public domain, but will provide protection for compilations of facts that meet the threshold for "originality." Such protection is considered to be "thin," as it does not extend to the underlying facts, applying only to their original selection or arrangement. In the European Union, database rights offer a more robust protection for compilations of data, but they also fall short when it comes to protecting the facts that make up such compilations.

Debates over ownership rights in data have been heating up. In Europe, policy makers have raised the possibility of creating *sui generis* ownership rights in data. In Canada, a recent court decision has raised the interesting question of whether facts and data should be treated differently in copyright law, offering a far more robust protection for data than for facts. In addition to these developments, Europe's new General Data Protection Regulation also appears to vest certain rights in data subjects through the newly introduced concept of data portability.

If data are capable of ownership, either through a *sui generis* right or copyright law, this raises important questions about how to strike a balance between the rights of data "owners" and the public interest in access to and reuse of data. This paper will explore the legal basis for claims of ownership of data, the extent of the public interest in access to and use of data, and the areas in which public policy development is required to address the changing needs of the data economy and society.

Introduction

Data have shifted from being a by-product of industrial, commercial, consumer and other activities to being a resource in their own right. The commercial value and importance of data is such that they have been referred to as the "new oil,"[1] although many are quick to point out that data are an infinitely renewable resource.[2] Unlike oil, the same data are also capable of being used by multiple actors and for multiple purposes. Data are the stock-in-trade of some of the world's largest corporations; they are also increasingly seen as the source of answers to society's problems. In this context, it is not surprising that data ownership issues have taken on great significance.

Ownership claims and legal skirmishes over rights to own or control data arise in a number of contexts (see Table 1, below). The announcement of a smart city collaboration between Waterfront Toronto and Sidewalk Labs, for example, has generated considerable discussion about who will own any data generated by the smart city.[3] In the United States, ongoing litigation between LinkedIn and companies that scrape LinkedIn's platform data raises a host of issues around ownership and control over publicly accessible platform data.[4] In the personal data context, there has

1 See e.g. "The world's most valuable resource is no longer oil, but data", *The Economist* (6 May 2017), online: <www.economist.com/news/leaders/21721656-data-economy-demands-new-approach-antitrust-rules-worlds-most-valuable-resource>. Kurtis McBride suggests that data is the new oil "[b]ecause it creates class warfare" (Kurtis McBride, "Monetizing Smart Cities: Framing the Debate" CIGI, Data Governance in the Digital Age Special Report, 28 March 2018, online: <www.cigionline.org/articles/monetizing-smart-city-data?utm_source=twitter&utm_medium=social&utm_campaign=data-series>.

2 See e.g. Bernard Marr, "Here's why data is not the new oil", *Forbes* (5 March 2018), online: <www.forbes.com/sites/bernardmarr/2018/03/05/heres-why-data-is-not-the-new-oil/#7951cd6c3aa9>; Adam Schlosser, "You may have heard data is the new oil. It's not", World Economic Forum (10 January 2018), online: <www.weforum.org/agenda/2018/01/data-is-not-the-new-oil/>.

3 See e.g. Teresa Scassa, "Who owns all the data collected by 'smart cities'?", *Toronto Star* (23 November 2017), online: <www.thestar.com/opinion/contributors/2017/11/23/who-owns-all-the-data-collected-by-smart-cities.html>; McBride, *supra* note 1; Bianca Wylie, "Civic Tech: The City of Toronto must remain a public platform", *The Torontoist* (10 January 2018), online: <https://torontoist.com/2018/01/civic-tech-keeping-city-public-platform/>.

4 See e.g. Kate Conger, "LinkedIn sues anonymous data scrapers", *TechCrunch* (15 August 2016), online: <https://techcrunch.com/2016/08/15/linkedin-sues-scrapers/>. See also *hiQ Labs Inc v LinkedIn Corp*, Case 3:17-cv-03301-EMC (ND Cal 2017), online: <https://regmedia.co.uk/2017/08/14/hiqlinkedintro.pdf>.

Table 1: Contexts in which Data Ownership Issues Arise

Data ownership can play a role in commercializing data: It is common for companies and organizations to seek to control the data they collect through their activities in order to commercialize them. An ownership right can support various techniques for control, including contracts/licensing and technological protection measures.
Data ownership can create monopolies: Regulators have already recognized the competition/antitrust issues that may arise from excessive concentrations of certain types of data in the hands of one or only a few companies.[5] Data ownership rights have a complicated relationship with competition law. This is, in part, because competition law creates some space for monopolies arising from intellectual property (IP) rights.[6]
Data ownership can have public dimensions: The open data movement involves governments making government data available for reuse under open licences. Underlying open data licences are claims to government ownership rights in the data. This is an example of where data ownership is used to pursue a particular public policy agenda.[7] Governments also have the option to not licence data as open data and, instead, to generate revenue through its sale, or to sell the data on a cost-recovery basis.
Data ownership may be challenging to locate: The recent Sidewalk Toronto smart cities project announced for Toronto has sharpened the focus on the importance of data ownership rights data within public-private partnerships involving the collection and/or generation of data. Who owns the data will have important implications not just for the ability to commercialize the data or to use or reuse it in the public interest; the location of ownership will also determine whether public or private sector data protection laws apply, whether the data can be sought through access to information requests, and even whether public-sector data localization laws will apply to prevent the storage of the data in another country.
Data ownership may play a role in privacy protection: Another dimension of ownership relates to privacy. The evolving European model of data protection will give individuals increased control over their data, including rights of erasure and data portability rights. These are quasi-ownership rights and may conflict with IP-style ownership rights in data.

been considerable discussion about individuals' ownership rights in their personal information,[8] including arguments that individuals should be able to monetize their personal data.[9] Indigenous data activists in Canada place the concepts of data ownership, control, access and possession at the heart of Indigenous data sovereignty.[10] In each of these examples, the value and importance of data is evident. Claims to ownership may be linked to the facilitation of commercial exploitation, but some claims also serve broader public interests.

The rights associated with ownership provide a powerful basis for control. A data owner can provide access to data or can restrict access partially or entirely. Data owners can impose conditions on access or use, including charging fees. Yet, if rights of ownership are recognized, the power to control information resources must have limits as well. IP laws create a balance of public and private interests; any data ownership rights must similarly include exceptions that are

5 See e.g. *Toronto Real Estate Board v Commissioner of Competition*, 2017 FCA 236 [*Toronto Real Estate Board*].

6 See e.g. *Competition Act*, RSC 1985, c C-34, s 79(5).

7 Note that there are other contexts in which data ownership rights may serve the public interest. For example, Giuseppina D'Agostino et al argue that "IP is an important legal tool for controlling and protecting medical data." See Giuseppina D'Agostino et al, "On the importance of intellectual property rights for e-science and the integrated health record" (2008) 14:2 Health Informatics Journal 95 at 98.

8 For example, see Jamie Lund, "Property Rights to Personal Information" (2011) 10:1 Northwestern J Technology & Intellectual Property 1–18; Jane B Baron, "Property as Control: The Case of Information" (2012) 18 Mich Telecomm & Tech L Rev 367–418.

9 For example, see Michelle Dennedy & Sagi Leizerov, "On monetizing personal information: A series", *The Privacy Advisor* (26 September 2017), online: <https://iapp.org/news/a/on-monetizing-personal-information-a-series/>; Michael Haupt, "Introducing Personal Data Exchanges & the Personal Data Economy", *Project 2030* (7 December 2016), online: <https://medium.com/project-2030/what-is-a-personal-data-exchange-256bcd5bf447>. This is discussed in greater detail in the subsection entitled "Data Ownership Right", below.

10 See e.g. First Nations Information Governance Centre (FNIGC), *Ownership, Control, Access and Possession (OCAP): The Path to First Nations Information Governance* (Ottawa: FNIGC, 2014) [FNIGC, OCAP], online: <https://fnigc.ca/sites/default/files/docs/ocap_path_to_fn_information_governance_en_final.pdf>. This is discussed in greater detail in the subsection entitled "Data Sovereignty", below.

appropriate for the public interest. The ability to access and use data is crucial to innovation, knowledge, transparency, accountability, expression and privacy. Further, not all rights or interests associated with data are ownership rights and, in some cases, different rights and interests may be layered with ownership rights.

Because the focus of this paper is on data ownership, the section entitled "Understanding Data" begins with an examination of what is meant by data. This is followed in the next section, entitled "Ownership under Existing Laws", by a discussion of the different legal regimes relevant to arguments about data ownership. The section, entitled "Proposals to Change Rules of Ownership", looks at potential alternatives to data ownership rules, including ownership rights in personal information, a *sui generis* data ownership right and data sovereignty claims. The paper concludes in the section entitled "Challenges with Data Ownership Rights" with an assessment of the existing frameworks for data ownership and an identification of those areas requiring policy attention.

Understanding Data

Understanding what is meant by the term data and by related terms, such as information and facts, is important given that, to the extent that there are any data ownership rights, they must attach to something about which there can be a clear consensus. Although the law has already grappled with concepts of ownership in relation to facts and information, the case law reveals inconsistencies in the understanding of these concepts.

Rob Kitchin describes three initial broad categories of data: data that are representative in nature, data that are implied and data that are derived.[11] Representative data are typically those that involve some kind of measurement, such as a person's age, the ambient temperature or the volume of traffic on a given road. Implied data are those read into an absence, such as inferences drawn about a person's voting preferences based on his or her online activity. Derived data are those that are "produced from other data."[12] These distinctions are important, and, as will be discussed below, early case law around data ownership evolved predominantly around representative data. In the context of big data[13] analytics, implied and derived data take on great importance and may include data profiles or predictive data. As will be seen in the discussion below, the distinction between representative data, on the one hand, and implied or derived data on the other is important to understanding some of the challenges with how the law addresses data ownership, in particular in the context of IP law.

Another important, although often overlooked feature of data is their non-neutrality. Kitchin writes that "data are in fact framed technically, economically, ethically, temporally, spatially and philosophically. Data do not exist independently of the ideas, instruments, practices, contexts, and knowledges used to generate, process and analyse them."[14] Data inherently reflect choices — choices about which data to collect (or to exclude) and what tools or parameters will be used in their collection. In the case of derived data, the data reflect the many choices that went into determining how they would be processed and for what ends. These choices reflect the human agency present in the creation of data.

The discussion of data also requires consideration of the concept of facts, which should not be conflated with data. Some see facts as the building blocks of data; in other words, facts occupy the role of the representative data referred to by Kitchin, whereas data are derived from them. In law, some courts have gone farther, attributing to facts an objective reality.[15] Some courts have treated facts as existing independently of those who record them, the latter being characterized as the "discoverers" of those facts.[16] Yet, to the extent

11 Rob Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures & their Consequences* (London: Sage, 2014) at 1.

12 *Ibid*.

13 Kitchin, *supra* note 11 at 68, notes that big data is often understood in relation to the concepts of volume, velocity and variety. Big data is huge in volume in that it exists in terabytes or petabytes; it is high in velocity in that it is created in real or near-real time; and it is available in a variety of formats. Other characteristics that may separate big data from small data are that it is "exhaustive in scope," "fine-grained in resolution," "relational in nature" and "flexible."

14 *Ibid* at 2.

15 In the Canadian case *R v Allen* [2006] AJ No 411 at para 11, the court, in noting that there can be no copyright in facts, bluntly stated: "Facts exist."

16 *Feist Publications Inc v Rural Telephone Service Co*, 499 US 340 at 347 (1991) [*Feist*].

that measurements are collected and recorded, for example, they inevitably reflect human choices. Further, not all facts are as simple as recorded measurements. In their book on the construction of facts in science, Bruno Latour and Steve Woolgar characterize scientific activity as "a fierce fight to construct reality" and consider scientific facts as "the set of statements considered too costly to modify constitute what is referred to as reality."[17]

Another term that sometimes clouds the discussion is information. Information can be understood as contextualized facts. Robert Losee notes that a common definition of information is "one or more statements or facts that are received by a human and that have some form of worth to the recipient."[18] While facts are raw, information is processed. For example, a reporter might collect different facts; her newspaper article contextualizes those facts and constitutes information. Copyright law recognizes this distinction. While it would protect her article as a "literary work," the facts embedded in the account would be considered in the public domain. Anyone would be free to take and reuse the facts, so long as they did not reproduce the original expression.[19] Yet, it can be difficult in some cases to distinguish between representative facts and information. This is not helped when the terms are sometimes used interchangeably.

Distinguishing between data and the works that contain them has taken on new significance in the big data era, as technology now permits the mining of data from all kinds of works, and those data have applications that go beyond the creation of new works. For example, a photograph is a copyright-protected work; facial recognition software can extract data about features from a photograph to create a faceprint that can be used in identifying the individual. From a copyright perspective, the question would be whether the faceprint is a reproduction of a substantial part of the protected expression in the photograph or whether all that has been extracted are public domain facts. Controversy currently swirls around the use of literary works to train algorithms for machine learning. Although these works must be copied in bulk for this purpose, they are not being used as "works," but rather as masses of data with uses that go beyond their original expression.[20] While it is not necessary to resolve these issues for the purposes of this paper, it is important to emphasize that the fact/data distinction is an important one for copyright law and is central to thinking about rights of data ownership.

Ownership under Existing Laws

As data take on an increasingly important economic role, the existing frameworks for IP ownership are called upon to provide some form of protection for rights in data. In practical terms, ownership rights are frequently asserted in data, although the nature, scope and robustness of these rights may be uncertain and contingent. In most cases, claims of ownership are based in copyright law or asserted under the laws of confidential information. In Europe, database protection laws also play a role. In addition to IP frameworks, data protection laws, which place individual control/consent at the heart of the protection of personal information, have created a context in which this element of control elides with the concept of ownership, supporting a growing discourse around individual ownership of personal information.

Copyright

Copyright law is an important source of contemporary claims to ownership rights in data. Where copyright arises, the author of a work is automatically protected (without need for registration) for the full term of protection.[21] Because the right is automatic and of considerable duration, and

17 Bruno Latour & Steve Woolgar, *Laboratory Life: The Construction of Scientific Facts* (Princeton, NJ: Princeton University Press, 1986) at 243.

18 Robert M Losee, "A Discipline Independent Definition of Information" (1997) 48:3 J American Society for Information Science 254 at 255.

19 *Maltz v Witterick* [2016] FCJ No 484, 2016 FC 524 at para 36 [*Maltz*].

20 See e.g. Christophe Geiger, Giancarlo Frosio & Oleksandr Bulayenko, "The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects" (2018) Center for International Intellectual Property Studies Research Paper No 2018-02, online: <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3160586>; European Commission, *Proposal for a Directive of the European Parliament and of the Council on Copyright in the Digital Single Market*, COM(2016) 593 final, 2016/0280 (COD), online: <https://ec.europa.eu/digital-single-market/en/news/proposal-directive-european-parliament-and-council-copyright-digital-single-market>.

21 In Canada, this is the life of the author plus an additional 50 years; in the United States and the European Union, it is life plus 70 years.

because it applies to all manner of expressive works, there are numerous exceptions to copyright protection, including fair dealing (or the related concept of fair use in the United States).

Copyright protection is available for literary, dramatic, artistic and musical works. These categories are defined broadly. For example, literary works have been found to include functional works, such as lists, menus, instructions and computer programs.[22] Artistic works include charts, plans, maps and graphic design, among others. The Copyright Act also recognizes that "compilations" consisting of multiple works of the same or different kinds, as well as compilations of fact, are also works in which copyright subsists.[23] The inclusivity of these broad categories is mitigated somewhat by the further requirement that a work be original.

Originality in Canadian copyright law has three main components. To be original, a work must not be a mere copy of another work;[24] it must also be the product of an exercise of skill and judgment.[25] Implicit in the skill and judgment requirement is the further requirement that there be a human author.[26] Although the Supreme Court of Canada (SCC) characterized this threshold for originality as lying somewhere between the British "sweat of the brow" standard and the US standard that requires a scintilla of creativity,[27] the reality is that the Canadian standard is comparable to those relied upon in the United States and the European Union, as well as in many other countries.[28]

No Copyright in Facts or Ideas

While the categories of works are interpreted broadly and inclusively, some subject matter is automatically excluded from protection under copyright law. Thus, there is no protection for ideas in the abstract — only for an original expression of ideas.[29] Even where an idea is expressed in an original way (for example, in a book or movie), the idea may still be used by another in an independent work without infringement, so long as the original expression of the idea is not copied.[30] Similarly, facts are not protectable under copyright law. Facts are considered to be the building blocks of knowledge and innovation, and copyright law has been careful not to permit their monopolization. The US Supreme Court has stated: "[A]ll facts — scientific, historical, biographical, and news of the day, may not be copyrighted and are part of the public domain available to every person."[31] According to the SCC, "copyright protection only extends to the expression of ideas as opposed to the underlying ideas or facts."[32] Justice Reed of the Federal Court described the proposition that facts are in the public domain as "trite law."[33] Thus, even where a work expresses facts in an original way, another party is free to extract those facts and use them in another work — providing they do not copy the original expression of those facts.[34]

Copyright in Compilations

As noted above, a work must be original to be protected by copyright. The originality standard is applied to works differently, depending on

22 Teresa Scassa, "Originality and Utilitarian Works: The Uneasy Relationship between Copyright Law and Unfair Competition" (2004) 1 University of Ottawa Technology LJ 51.

23 Section 2 of the *Copyright Act*, RSC 1985, c C-42, defines "compilation" as "(a) a work resulting from the selection or arrangement of literary, dramatic, musical or artistic works or parts thereof, or (b) a work resulting from the selection or arrangement of data."

24 *CCH Canadian Ltd v Law Society of Upper Canada*, [2004] 1 SCR 339, 2004 SCC 13 (CanLII) at para 25 [*CCH Canadian Ltd*], online: <http://canlii.ca/t/1glp0>.

25 *Ibid* at para 16. Chief Justice McLachlin, writing for the court, stated: "By skill, I mean the use of one's knowledge, developed aptitude or practised ability in producing the work. By judgment, I mean the use of one's capacity for discernment or ability to form an opinion or evaluation by comparing different possible options in producing the work. This exercise of skill and judgment will necessarily involve intellectual effort."

26 See *CCH Canadian Ltd*, *supra* note 24 at para 15. The issue of human authorship has arisen in the context of compilations of facts. For example, in the Australian case of *Telstra Corporation Limited v Phone Directories Company Pty Ltd*, [2010] FCA 44 at paras 90–91, 333, the court found it impossible to identify a human author for the telephone directory compiled through automated processes. In *Geophysical Service Inc v Encana Corp*, 2016 ABQB 230, 38 Alta LR (6th) 48, aff'd 2017 ABCA 125, leave to appeal denied 2017 CanLII 80435 (SCC) [*Geophysical Service Inc*], the court rejected arguments that the collection of seismic data, collected through the use of sophisticated technology, lacked a human author. Justice Eidsvik found that "[h]uman input is involved continuously through the acquisition stage, like creating a sound recording" (*ibid* at para 90). She noted that whether there was human authorship or not was an issue to be decided on the facts of a particular case.

27 *Ibid* at paras 15–16.

28 See e.g. Daniel Gervais, "Canadian Copyright Law Post-CCH" (2004) 18 IPJ 131. The standard in the United Kingdom may lean more toward recognizing industrious creation as a basis for copyright protection than does the Canadian standard, as noted by the SCC in *CCH Canadian Ltd*, *supra* note 24.

29 David Vaver, *Intellectual Property Law: Copyright, Patents, Trade-marks*, 2nd ed (Toronto: Irwin Law, 2011) at 59–60.

30 See e.g. *Maltz*, *supra* note 19.

31 *Feist Publications Inc v Rural Telephone Service Co*, 499 US 340, 111 S Ct 1282 at 348 (1991) [*Feist*].

32 *CCH Canadian Ltd*, *supra* note 24 at para 15.

33 *Hager v ECW Press Ltd*, [1999] 2 FC 287, 1998 CanLII 9115 (FC) at para 44, online: <http://canlii.ca/t/1hcj4>.

34 *Ibid*; *Maltz*, *supra* note 19.

their nature. In the case of compilations, it is the authorial contribution of the person who created the compilation that is assessed, since this is what copyright seeks to protect. The author of a compilation is not generally the author of the individual works of which it is composed (for example, the creator of an anthology of stories is not typically the author of the stories it contains). The originality of a compilation, therefore, lies in the work of its author in selecting and arranging the different elements it contains.[35] If there is an original selection or arrangement, this can give rise to copyright protection. The extent of the protection is limited to the original selection or arrangement — it does not extend to the individual elements that make up the compilation. Thus, copyright in a compilation of facts is infringed if someone takes a substantial part of the author's original *selection* or *arrangement* of those facts.[36]

In the case of compilations of facts, courts have found that telephone directories are not sufficiently original to give rise to copyright protection, since neither the selection of the individual facts contained therein, nor their arrangement (in alphabetical order) are original.[37] Even so, the threshold for originality is not particularly high. For example, thematic telephone directories may be protected by copyright if the theme reflects an original selection.[38] Nevertheless, compilations of fact present many challenges when it comes to copyright. "Whole universe" sets of fact may not reflect an original selection; similarly, where facts are arranged according to industry norms or standards, the compilation may lack originality.[39] A data set that is constantly growing (for example, streamed sensor data) may similarly be incapable of being a compilation since there is never a completed work. Even if a selection or arrangement is original, the principle that facts are in the public domain means that only the original selection or arrangement of the compilation will be protected; anyone who extracts facts from the compilation using an independent selection and arrangement of those facts has not infringed copyright. This has led the US Supreme Court to describe copyright in factual compilations as "thin."[40]

The Database Industry and Database Rights

Before considering copyright issues in relation to data, it is important to reflect on the historical context in which the jurisprudence around copyright in facts has evolved. At one point, courts in North America and the United Kingdom showed a tendency toward recognizing the considerable effort that often went into compiling a collection of facts. This sweat-of-the-brow doctrine reflected concerns about unfair competition and recognized that a competitor of a compiler of facts could profit from the effort involved in creating the compilation to the detriment of the party that had invested the time, energy and money necessary to create it in the first place.[41] To a large extent, this doctrine arose and was developed in a pre-digital era, when collection and compilation was often a painstaking and laborious enterprise.

The case law that disrupted the trend toward sweat of the brow arose in the 1990s, at a time when digital technologies were emerging, creating greater opportunities for the automated collection and processing of facts. The signature cases in the United States and Canada that dismissed sweat of the brow as a basis for finding copyright in a compilation of facts involved telephone directories. Their backstories also revealed concerns about monopolies,[42] as well as the potential for stifling innovation in spin-off products that

35 *Robertson v Thomson Corp*, [2006] 2 SCR 363, 2006 SCC 43 (CanLII), online: <http://canlii.ca/t/1pqw1>; *Tele-Direct (Publications) Inc v American Business Information, Inc* (CA), [1998] 2 FC 22 [*Tele-Direct*].

36 *CCH Canadian Ltd*, *supra* note 24 at para 33.

37 In Canada, see e.g. *Tele-Direct*, *supra* note 35. The US Supreme Court reached a similar conclusion in *Feist*, *supra* note 31.

38 See e.g. *ITAL-Press Ltd v Sicoli*, [1999] FCJ No 837 (FC); *B & S Publications Inc v Max-Contacts Inc*, [2001] AJ No 143. In the United States, see *Key Publications Inc v Chinatown Today Publishing Enterprises Inc*, 945 F (2d) 509 (2nd Cir 1991).

39 See e.g. *Tele-Direct*, *supra* note 35. Industry standards, which may be relevant, but are not determinative of the originality enquiry, are considered in *Harmony Consulting Ltd v GA Foss Transport Ltd*, 2011 FC 340 at paras 34, 39, 65, 77, 182–188, 92 CPR (4th) 6, aff'd 2012 FCA 226 at paras 37–38, 107 CPR (4th) 1; *Geophysical Service Inc*, *supra* note 26 at para 105. In *Toronto Real Estate Board*, *supra* note 5 at para 194, the Federal Court of Appeal found that the process by which the Toronto Real Estate Board compiled its data was a "mechanical exercise" and that the compilation therefore lacked originality.

40 *Feist*, *supra* note 31 at 349.

41 For a discussion of the historical evolution of the approach to copyright in facts, see Miriam Bitton, "Feist, facts and functions: historical perspective" in Robert F Brauneis, ed, *Intellectual Property Protection of Fact-based Works: Copyright and Its Alternatives* (Cheltenham, UK: Edward Elgar, 2009).

42 At the time, telephone service providers had privileged access to subscriber data and often also had statutory obligations to produce telephone directories for public use. In this context, copyright protection of the resulting directories would actually enhance an already existing monopoly without providing any additional incentive to collect the facts or produce the directories.

might be developed using facts. Courts were applying copyright law principles within a very particular economic and technological context. For example, the US Supreme Court, in explaining the rationale for facts remaining in the public domain, used the example of a census taker as a collector of facts, ruling that such individuals "copy these figures from the world around them."[43] The model was of individualized, painstaking recording of representative facts.

Following the decision in *Feist*, concerns over the vulnerability to unfair competition of producers of compilations of data led to calls on both sides of the Atlantic for new legislation to specifically protect databases. In Canada and the United States, the choice was made to do nothing — essentially allowing the database industry to emerge and evolve in a context in which protections other than copyright law would also play a role.[44] In Europe, however, the decision was made to create a *sui generis* database right.[45] The European Database Directive required member states to enact laws that would protect the investment that went into the creation of databases. At the heart of such protection, however, was not a property right in data, but rather the protection of the investment made in structuring the data — in other words, in creating the database.[46] As Thomas Hoeren notes, any investment in the creation of the *data* themselves is not protected.[47]

Copyright in Data

The historical context around the exclusion of facts from copyright protection and the limited protection available for compilations of fact is important because in the big data era, this context has markedly changed, and there is reason to believe that case law is also beginning to shift as well. Massive advances in digital technologies mean that the collection of data is constant and ubiquitous — it is often not the painstaking, resource-intensive or laborious process that gave rise to sweat-of-the-brow approaches, nor is it the automated but bounded collection and compiling of facts that made up the more recent era. Today, data are not just collected ubiquitously and continuously, they are processed, analyzed and stored in increasingly complex ways; they are also used to generate new data in the form of profiles, predictions and analytics. In this context, a focus on the originality of selection or arrangement is quickly overwhelmed by the human and technological processes that underlie the big data economy.

Within this evolving context, the importance of the definitions considered at the outset of this paper is evident. Copyright law clearly does not protect facts — facts are considered to be in the public domain. However, it is less clear that data are not protected. This view finds some support in the international treaty context. Article 2(8) of the Berne Convention provides that "[t]he protection of this Convention shall not apply to news of the day *or to miscellaneous facts having the character of mere items of press information.*"[48] This suggests that while certain types of facts are in the public domain, others may not be. The distinction between "mere" facts and the more complex concept of data may also be evident in the wording of the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement), which provides in article 10(2) that:

> Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations shall be protected as such. Such protection, *which shall not extend to the data or material itself, shall be without prejudice to any copyright subsisting in the data or material itself.*[49]

43 *Feist*, *supra* note 31 at 347.

44 For example, contractual terms of use are commonplace.

45 EC, *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases*, [1996] OJ, L 77/20, online: <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31996L0009:EN:HTML> [*European Database Directive*].

46 Thomas Hoeren, "Big Data and the Ownership in Data: Recent Developments in Europe" (2014) 12 EIPR 751 at 752.

47 *Ibid*, citing *British Horseracing Board Ltd v William Hill Organization Ltd* (C-203/02) [2004] ECR I-10415; [2005] 1 CMLR 15 at para 33.

48 *Berne Convention for the Protection of Literary and Artistic Works*, 9 September 1886, 25 UST 1341, 828 UNTS 221 (amended 28 September 1979) [emphasis added], online: <www.wipo.int/wipolex/en/treaties/text.jsp?file_id=283698>.

49 *Agreement Establishing the World Trade Organization, Annex 1C: Agreement on Trade-Related Aspects of Intellectual Property Rights*, 15 April 1994, 1869 UNTS 299 art 10(2) (entered into force 1 January 1995), online: <www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm> [*TRIPS Agreement*] [emphasis added]. Also note that the *North American Free Trade Agreement*, 17 December 1992, Can TS 1994 No 2, 32 ILM 289, 605 (entered into force 1 January 1994) (Ottawa: Supply & Services, 1993), online: <www.nafta-sec-alena.org/Home/Texts-of-the-Agreement/North-American-Free-Trade-Agreement?mvid=1&secid=b6e715c1-ec07-4c96-b18e-d762b2ebe511#A1705>, contains similar wording in article 1705(1).

While on the one hand, this might be read as acknowledging the potential for separate copyright to exist in the elements of a compilation (such as individual stories in an anthology), the reference to "any copyright subsisting in the data" leaves open the possibility that, while copyright in a *compilation* of data does not extend to the underlying data, it is possible that a separate copyright might subsist in the data.

There are numerous instances in copyright jurisprudence where courts have equivocated about the nature of facts. A pair of US cases is illustrative. In *Castle Rock Entertainment, Inc. v. Carol Publishing Group, Inc.*,[50] the court had to determine whether well-known trivia about episodes of the popular television series *Seinfeld* were facts (and therefore in the public domain and available to the creator of a trivia game based on the series), or a substantial part of the copyright-protected work that was the series. The court characterized the trivia as "expression" created by the authors of the series.[51] It also distinguished between real facts (such as the names of the actors who play the characters) and fictionalized facts (those arising from events in the series). In *Warner Bros. Entertainment Inc. v. RDR Books*,[52] a court reached a similar conclusion about the defendant's lexicon, which was derived from the Harry Potter series of novels. According to the court, "[e]ven if expression is or can be used in its 'factual capacity,' it does not follow that expression thereby takes on the status of fact and loses its copyrightability."[53] In both of these cases, the courts recognized a degree of authorship in the creative facts, making them an integral part of the expressive work as a whole.

The case of facts arising from fictional works might be considered particular, but the argument that other types of data are authored arises in other circumstances as well.[54] In *New York Mercantile Exchange Inc. v. Intercontinental Exchange Inc.*,[55] the US Court of Appeal for the Second Circuit considered claims that the defendant's copying of settlement prices generated by the plaintiff's algorithm amounted to copyright infringement. Ultimately, the court characterized the issue as one of determining "the line between creation and discovery,"[56] in other words, whether the plaintiff was the author of the settlement prices or merely their discoverer. It considered this to be a "close question."

Similarly, in *BanxCorp v. Costco Wholesale Corp.*,[57] the court considered arguments that calculated percentages were original and therefore protected by copyright. The court seemed to accept the possibility that some facts could be authored, but observed that "when confronted with raw data that have been converted into a final value through the use of a formula, courts should put significant weight on the degree of consensus and objectivity that attaches to the formula to determine whether the final value is fundamentally a 'fact.'"[58] The court attempted to distinguish between facts that merited protection and those that did not, stating:

> If the data purports to represent actual objective prices of actual things in the world — the actual price of an actual settlement contract on a particular day — it is an unprotectable fact; if the data purports to represent an estimated price of a kind of idealized object — for instance, what a hypothetical, mint condition 2003 Ford Taurus with approximately 60,000 miles might be worth — then the hypothetical price may be eligible for some form of copyright protection in the right circumstances.[59]

These decisions leave open the possibility that some data might be entitled to copyright protection.

50 *Castle Rock Entertainment, Inc v Carol Publishing Group, Inc*, 150 F (3d) 132 (2nd Cir 1998).

51 *Ibid* at 139.

52 *Warner Bros Entertainment Inc v RDR Books*, 575 F Supp (2d) 513 (SDNY 2008).

53 *Ibid* at 536.

54 For a discussion of claims to copyright in bus timetable data, see Teresa Scassa, "Copyright Reform and Fact-Based Works" in M Geist, ed, *From "Radical Extremism" to "Balanced Copyright": Canadian Copyright and the Digital Agenda* (Toronto: Irwin Law, 2010) 571.

55 *NY Mercantile Exch, Inc v Intercontinental Exchange, Inc*, 497 F (3d) 109 (2d Cir 2006) [*NY Mercantile Exch*].

56 *Ibid* at 114. See also *RBC Nice Bearings, Inc v Peer Bearing Co*, 676 F Supp (2d) 9, 21 (D Conn 2009), which found that data derived from a series of calculations carried out by the plaintiffs still fell within the category of facts.

57 *BanxCorp v Costco Wholesale Corp*, 978 F Supp (2d) 280 (SDNY 2013).

58 *Ibid* at 300.

59 *Ibid* at 301. The court in this passage is referencing CCC *Information Services, Inc v Maclean Hunter Market Reports, Inc*, 44 F (3d) 61 (2d Cir 1994).

The merger doctrine was ultimately relied upon in *Mercantile Exchange* to avoid a finding of copyrightability. This same doctrine was also invoked in *Banxcorp*. The merger doctrine is interesting in this context and may have a role to play in considering whether data are copyrightable. The merger doctrine, which has been considered, although not expressly adopted by courts in Canada,[60] is based on the principle that there can be no copyright in either facts or ideas, but only in their original expression. Where the expression of a fact or an idea merges with that fact or idea, (for example, where there is only one or a very limited number of ways to express it), there can be no copyright protection since the practical result of any such protection would be to give a monopoly over the fact or idea. For example, the court in *NY Mercantile Exchange* ruled that the merger doctrine applied to the settlement prices, stating that "[b]ecause any settlement price for a particular futures contract would be determined based on the same underlying market facts, any dissension would be exceptionally narrow."[61]

The role of the merger doctrine in determining whether implied or derived data can be protected by copyright is intriguing. *NY Mercantile Exchange* suggests that derived data can be authored and therefore original. However, such data, to the extent that they represent the idea behind the analytics that led to their creation, reflect a merger of idea and expression. If this is the case, then it would seem that derived data must necessarily remain in the public domain, except where there is no merger between idea and expression. The challenge will be in determining when no merger occurs.

The merger doctrine aims to avoid giving a monopoly to one party over an idea, thus preventing others from engaging with that idea expressed in its simplest form. However, an expansive application of the merger doctrine might stifle innovation if innovators lose all advantage in generating new works because others are free to copy and use them. However, in *NY Mercantile Exchange*, the court noted that the plaintiff produced the settlement figures in order to facilitate futures trading and that this was the only incentive needed to produce them. From a public policy point of view, copyright protection would merely block others from using the figures rather than support the continued creation of the figures by the plaintiff. The court, thus, also relied on the *purpose* of copyright law as a means to determine whether copyright protection was warranted.

In *Geophysical Service Inc. v. Encana Corp*,[62] a judge of the Alberta Court of Queen's Bench made an interesting distinction between public domain facts and copyrightable data. The case involved, among other things, claims of copyright in the plaintiff's seismic data about the ocean floor. The data were collected through a process that required considerable skill, as well as time and resources. Justice Eidsvik divided the data into two categories: field data and processed data. The field data were the raw data collected using the plaintiff's tools and technology. The processed data were defined as "any product derived, generated or created from the data, including, but not limited to any and all processed and reprocessed data, interpretations, maps or analyses, regardless of the form or medium on which it is displayed or stored."[63] Justice Eidsvik found that the field data were a compilation, while the seismic data were either a compilation or an artistic work, depending on their representation. She found that both categories of data met the threshold for originality, as each required considerable skill and judgment for their creation. Although technology played a considerable role in the collection and generation of the data, there was still sufficient human authorship to give rise to copyright. Justice Eidsvik concluded that the data were "an expression of GSI's [Geophysical Services Inc.'s] views of what the image of the subsurface of the surveyed areas represents."[64] She considered the uncopyrightable facts to exist in the features of the sub-ocean landscape; by contrast, the data were the details collected about

60 See e.g. *Delrina Corp v Triolet Systems Inc* (2002), 17 CPR (4th) 289 at paras 48–52 (Ont CA), leave to appeal refused, [2002] SCCA No 189, (2002), 305 NR 398 [*Delrina*]. Although no Canadian court has expressly adopted the merger principle, it is consistent with Canadian copyright law principles and has been referenced in many cases. See e.g. *Delrina* (*ibid*); *Distrimedic Inc v Dispill Inc*, [2013] FCJ No 1093 at para 323; *Red Label Vacations Inc v 411 Travel Buys Ltd*, 2015 FC 18, 473 FTR 38 at para 98.

61 *NY Mercantile Exch*, *supra* note 55 at 118. The merger doctrine was also applied to disqualify copyright protection for facts generated through a company's own calculations in *BanxCorp*, *supra* note 57. The court also held that, considered individually, the facts (calculated percentages) were too short to constitute expressive works.

62 *Geophysical Service Inc, supra* note 26.

63 *Ibid* at para 58, quoting from the Association of Professional Engineers, Geologists and Geoscientists of Alberta, *Guideline for Ethical Use of Geophysical Data*, vol 1, online: <www.apega.ca/assets/PDFs/geophysical-data.pdf>.

64 *Ibid* at para 97.

those features by the plaintiff. Anyone else who tried to collect the same underwater seismic data from the same locations would be unlikely to be able to completely replicate the data. Although Justice Eidsvik's conclusion is that "the raw or field seismic data is an original literary compilation work and the processed data is both an original literary compilation work and an artistic compilation work in the scientific domain,"[65] she seemed to recognize copyright in the data themselves rather than just in the overall compilations of data.[66]

It is generally understood that copyright protection requires a human author. Works that are created by automated processes in which human authorship is lacking cannot, therefore, be copyright protected.[67] This has raised concerns that the output of artificial intelligence (AI) processes will not be capable of copyright protection.[68] While these debates have relevance in the context of AI — and by extension to contexts where data are generated through AI — they do not, by any means, determine the issue of whether some data are capable of protection under copyright law. As noted above, courts in Canada and the United States have found sufficient authorship in data generated either by non-AI algorithms or by complex processes such as those used in the collection of underwater seismic data.

The upshot of this case law is that, notwithstanding the general principle that there is no copyright in facts, data may be treated differently in both Canada and the United States. This is particularly the case with implied or derived data, as opposed to merely representative data. Not only is it open to a court to conclude that the data themselves meet the threshold for originality; copyright protection is also available to the overall compilation of data where there is an original selection or arrangement. However, in copyright law, the merger doctrine and the public domain nature of facts and ideas help ensure that the public interest is part of the balance, and that copyright in data will always be contingent on the particular nature of the data and its expression. These doctrines and principles that limit copyright in data might be seen by some as disincentives to innovate, yet they also serve important public policy goals that support innovation and competition. There remains considerable and valuable flexibility in copyright law and its core principles.

A major part of the balancing of public and private interests in copyright law occurs through the fair dealing ("fair use" in the United States) provisions of the Copyright Act. The fair dealing exceptions to copyright infringement have been characterized as "users' rights."[69] In Canada, to qualify as fair dealing, a dealing with a work must be for one of the purposes set out in the statute[70] and it must be "fair."[71] Fair dealing purposes are typically ones that serve a public interest and can include research, private study, criticism or comment, education, parody or satire.[72] To the extent that copyright is recognized in either or both data and compilations of data, fair dealing rights would apply. This would create space for the equitable use of the data or compilations, for example, by researchers or by those seeking to challenge or criticize certain data or data sets.

Technological Protection Measures

The World Intellectual Property Organization (WIPO) Copyright Treaty (WCT)[73] and the WIPO Performances and Phonograms Treaty (WPPT)[74] both require signatory states to enact provisions to bolster the protection available to copyright owners who use technological protection measures (TPMs) to protect their content. For example, the WCT provides: "11. Contracting Parties shall provide adequate legal protection and effective legal remedies against the circumvention of effective technological measures that are used by authors in connection

65 *Ibid* at para 115.

66 *Ibid* at para 85; in *Geophysical Service Inc, supra* note 26, Justice Eidsvik stated, "the seismic data is an 'original' work."

67 See e.g. *Telstra Corporation Limited v Phone Directories Company Pty Ltd*, [2010] FCA 44, where the High Court of Australia found that there was insufficient human authorship in the automated process for creating telephone directories.

68 See e.g. Kalin Hristov, "Artificial Intelligence and the Copyright Dilemma" (2017) 57 IDEA 431; Shlomit Yanisky-Ravid, "Generating Rembrandt: Artificial Intelligence, Copyright, and Accountability in the 3A Era — The Human-like Authors Are Already Here — A New Model" (2017) Mich St L Rev 659.

69 *CCH Canadian Ltd*, *supra* note 24 at para 12.

70 *Copyright Act*, *supra* note 23, ss 29, 29.1, 29.2. These purposes include research, private study, criticism, comment, news reporting, education, parody and satire.

71 *CCH Canadian Ltd*, *supra* note 24 at para 50. The court goes on to identify six criteria for assessing the fairness of any dealing.

72 *Copyright Act*, *supra* note 23.

73 *WCT*, 20 December 1996, TRT/WCT/001 (entered into force 6 March 2002).

74 *WPPT*, 20 December 1996, TRT/WPPT/001 (entered into force 20 May 2002).

with the exercise of their rights under this Treaty or the Berne Convention and that restrict acts, in respect of their works, which are not authorized by the authors concerned or permitted by law."[75]

Canada did not implement this part of the WCT or the WPPT until 2012, when amendments to the Canadian Copyright Act added new provisions designed to protect against the circumvention of TPMs that are put in place to protect digital copyright-protected works. Section 41.1(1) of the Canadian Copyright Act provides that "[n]o person shall (a) circumvent a technological protection measure within the meaning of paragraph (a) of the definition *technological protection measure* in section 41."[76] Section 41 defines a TPM as "any effective technology, device or component that, in the ordinary course of its operation, (a) controls access to a work, to a performer's performance fixed in a sound recording or to a sound recording and whose use is authorized by the copyright owner; or (b) restricts the doing — with respect to a work, to a performer's performance fixed in a sound recording or to a sound recording — of any act referred to in section 3, 15 or 18 and any act for which remuneration is payable under section 19."[77]

Thus, to be protected under this scheme, a TPM must be "effective" and it may either control access to a work or restrict certain uses of it. The equivalent provision in the US Copyright Act provides: "No person shall circumvent a technological measure that effectively controls access to a work protected under this title."[78] In what little case law there has been to date in Canada, the definition of an "effective technological protection measure" has received a broad interpretation. In *Nintendo of America Inc. v. King*, the court stated: "The open-ended language of this definition reflects Parliament's intention to empower copyright owners to protect their business models with any technological tool at their disposal."[79] In the case of TPMs that provide access control, such as passwords, the court noted that these "do not need to employ any barrier to copying in order to be 'effective.'"[80]

While the motivation for such provisions may have been to provide additional protection in the battle against illegal copying of movies, music and other digital mass market works, they could have a significant effect on the scope of protection for data in copyright law. Since a compilation of data can be a copyright-protected work, the addition of TPMs to such compilations will provide a new level of protection for the data. This is particularly so in Canada, where, unlike in other countries, such as the United States,[81] fair dealing is not a defence to circumvention. The result is that anyone who circumvents a TPM in Canada in order to access a compilation of data with the goal to extract the public domain facts contained therein, or any data that is unprotected under the merger doctrine, might not be violating copyright in the compilation, but could be liable for circumvention. This is a troubling result, as it *de facto* extends the degree of protection available for compilations of data to the underlying data themselves, without any room to apply the balancing principles found in the merger doctrine or in fair dealing exceptions.

Confidential Information/ Trade Secrets

Information may be protected as confidential in certain circumstances.[82] These are described in article 39(2) of the TRIPS Agreement. Information can be protected as confidential information if it "(a) is secret in the sense that it is not, as a body, or in the precise configuration and assembly of components, generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question; (b) has commercial value because it is secret; and (c) has been subject to reasonable steps

75 A similar provision is found in the WPPT, art 18.

76 *Copyright Act*, *supra* note 23, s 41.1(1) [emphasis added].

77 *Ibid*, s 41.

78 *US Copyright Act*, 17 USC § 1201(a)(1)(A).

79 *Nintendo of America Inc v King*, 2017 FC 246 at para 73.

80 *Ibid* at para 84.

81 In the United States, courts have given a more nuanced interpretation to TPMs. This may, in part, be due to the wording of the US Copyright Act, *supra* note 76, which expressly addresses concerns over fair use. Section 1201(c)(1) provides that "[n]othing in this section shall affect rights, remedies, limitations, or defenses to copyright infringement, including fair use, under this title."

82 The broad category of confidential information includes trade secrets, which are typically of a more industrial nature, such as formulae or industrial processes.

under the circumstances, by the person lawfully in control of the information, to keep it secret."[83]

Canadian law is consistent with these conditions. The requirement that information not be "generally known" or "readily accessible" leaves room for the sharing of confidential information with, for example, prospective clients or investors. The "reasonable steps" component would require that any such sharing be done under the protection of confidentiality agreements and any other necessary security arrangements.

The protection available in law for confidential information is not the equivalent of an ownership right. In *R. v. Stewart*,[84] the SCC considered whether data could be "owned" for the purposes of determining whether it could be stolen within the meaning of the Criminal Code[85] provisions for theft. The court found that data could be misappropriated without the "owner" being deprived of the data — for example, where data is memorized or copied. As a result, they declined to find that it was property for the purposes of the Criminal Code. While the court did not consider whether it might be property in other contexts, it seems unlikely. In Canada, the law of confidential information is based on a number of different areas of law, including contract law, the law of fiduciary relationships, equity and tort, all of which tend to emphasize relationships between individuals. Rather than providing an IP-type monopoly, the law of confidential information tends to encourage fair competition and ethical behaviour, in part by "promoting, protecting and enforcing relationships founded on trust and confidence."[86] Courts have recognized duties of confidence in fiduciary relationships,[87] in the relationship between a company and its senior officers,[88] and in relationships between employees and their employers.[89] Obligations of confidence within these relationships can also be bolstered by contractual agreements and undertakings.

The non-proprietary nature of confidential information can be seen in the fact that its value typically lies in its confidentiality, and not in the information itself. Once confidentiality is lost, the information is often rendered valueless. The law of confidential information can be used to protect subject matter that would otherwise be unprotectable by other areas of IP law. For example, ideas cannot be protected under copyright law, but an idea can be shared under a confidentiality agreement. The law of confidential information can also be used to give a level of protection not available under other areas of IP law. Copyright law might protect the expression of a secret formula, but its commercial value is better protected by confidentiality. When it comes to inventions, while a particular invention might be patentable, a patent application requires full disclosure. An inventor might choose instead to protect his or her invention by maintaining its confidentiality in order to avoid sharing its details with the public (and with competitors).

Where federal and provincial laws require the submission of data to government for regulatory or other purposes, that data is typically protected as confidential commercial information. Access to information laws at the federal and provincial levels in Canada contain exceptions to the general right of access to information in the hands of government when it is the confidential commercial information of third parties.[90]

Relying on the law of confidential information to protect data may have a number of benefits, including the breadth of subject matter protected, the potentially infinite duration of protection and the relative paucity of public interest exceptions permitting access or reuse. Nevertheless, not all data can be protected as confidential information. For example, some data are necessarily broadly shared or are even publicly accessible.[91] Further, under the

83 *TRIPS Agreement*, *supra* note 49, art 39(2).

84 *R v Stewart*, [1988] 1 SCR 963, 1988 CanLII 86 (SCC) [*Stewart*].

85 *Criminal Code*, RSC 1985, c C-46.

86 Greg Hagen et al, *Canadian Intellectual Property Law: Cases and Materials*, 2nd ed (Toronto: Emond, 2017) at 579.

87 See e.g. *Lac Minerals Ltd v International Corona Resources Ltd*, [1989] 2 SCR 574, 1989 CanLII 34 (SCC), online: <http://canlii.ca/t/1ft3w>.

88 See e.g. *Canadian Aero Service Ltd v O'Malley*, 1973 CanLII 23 (SCC), [1974] SCR 592.

89 See e.g. *RL Crain Ltd v RW Ashton & Ashton Press Manufacturing Co* (1949), [1950] OR 62, [1950] 1 DLR 601 (ONCA); *Imperial Sheet Metal Ltd v Landry*, 2007 NBCA 51.

90 See e.g. *Access to Information Act*, RSC 1985, c A-1, ss 20(1)(a), (b) [*ATIA*]; *Freedom of Information and Protection of Privacy Act*, RSO 1990, c F 31, s 17(1). Note that at the federal level in Canada, this is an absolute exception to the requirement of disclosure; there is a discretionary element in Ontario. These exceptions can also extend to "commercially sensitive information" (see e.g. *ATIA* [*ibid*], ss 20[1][c], [d]).

91 See, for example, the discussion of publicly accessible platform data in Teresa Scassa, "Sharing Data in the Platform Economy: A Public Interest Argument for Access to Platform Data" (2017) 50:4 UBC L Rev 1017 [Scassa, "Sharing Data"].

law of confidential information, once confidentiality is lost, protection is effectively at an end.

Personal Information

Personal information is generally not capable of ownership — at least not by the persons to whom it pertains — although in recent privacy discourse, it is increasingly common to hear references to individuals "owning" their personal information. Certainly, the consent model of data protection is designed to give individuals a degree of control over their personal information. Recent developments under the EU General Data Protection Regulation[92] around the right of data portability, for example, also seem to lean toward quasi-ownership rights. Nevertheless, the control provided under data protection laws falls short of ownership, and even data portability is a carefully constrained type of control.

In Canada, the SCC's decision in *McInerney v. McDonald* [93] sheds some light on how the law construes the relationship between the data subjects and their personal information. At issue in *McInerney* was a patient's relationship to her medical records. More specifically, in the words of Justice La Forest, it was "whether in the absence of legislation a patient is entitled to inspect and obtain copies of his or her medical records upon request."[94] One of the theories considered, and ultimately rejected, by the court was that a patient owned their personal medical information. Instead, the court found that the "physician, institution or clinic compiling the medical records owns the physical records."[95] A patient shares information with a doctor in the context of a relationship of trust. The court went on to state that "[w]hile the doctor is the owner of the actual record, the information is to be used by the physician for the benefit of the patient. The confiding of the information to the physician for medical purposes gives rise to an expectation that the patient's interest in and control of the information will continue."[96] Thus, the court recognized the property right of a doctor in the physical medical record[97] and an "interest" on the part of the patient amounting to a degree of control over the information. The court explicitly considered and rejected arguments that this interest was a property interest. Justice La Forest stated: "I find it unnecessary to reify the patient's interest in his or her medical records."[98] The patient's interest was characterized as a right of access to the information.

Although *McInerney* dealt with personal health information, there is no reason to expect that a Canadian court's decision would be different with respect to other types of personal information. Indeed, public and private sector data protection laws in Canada tend to follow the model sketched out in *McInerney*. Those who collect personal information from individuals owe certain duties to the individual. These include complying with the norms for collecting, using, disclosing, storing and disposing of the information, as well as managing the information in accordance with the consent provided by the individual. Individuals have a right to access and to correct their personal information, but these rights stop short of ownership rights.

Proposals to Change Rules of Ownership

The principles regarding the ownership of data that are set out above are well-established in Canada and are generally consistent with the principles in place among its major trading partners. Nevertheless, with the rise in the economic importance of data, there has been a corresponding interest in reopening the debate and discussion around data ownership rights. In this section, the author considers three of these discussions: a personal data

92 *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*, [2016] OJ, L 119.

93 *McInerney v MacDonald*, [1992] 2 SCR 138, 1992 CanLII 57 (SCC), online: <http://canlii.ca/t/1fsbl>.

94 *Ibid* at para 1.

95 *Ibid* at para 14.

96 *Ibid* at para 22.

97 Note that the court talked of property rights in the physical records; it did not address the issue of IP rights in the contents of the file.

98 *Ibid* at para 25.

ownership right, a data ownership right for the commercial sector and data sovereignty.

Ownership of Personal Data

There are frequent references in popular discourse to individuals owning their personal information, although as noted earlier, this is not consistent with the current state of the law.[99] Nevertheless, as the commercial value of personal information has risen, some have argued that individuals should be able to commercialize their own personal information by exchanging its use for compensation. Some of the earliest examples of this appear in the business literature in the late 1990s, with calls for the establishment of infomediaries to facilitate such exchanges.[100] Other scholars have suggested that the concept of property rights should extend to personal information.[101] Kenneth Laudon, for example, suggested that privacy protection required a market approach: "if privacy is to be taken seriously as a public value, the solution is to rely on more powerful and less wasteful mechanisms, like markets, to reduce privacy invasion."[102]

More contemporary authors have also suggested different ways in which market-based solutions to privacy problems might operate.[103] This is referred to as the "personal data economy."[104] In recent years, start-ups have appeared that offer personal data management services to internet users. These range from companies that pay users for their personal information to those that allow users to pay supplementary fees in order to avoid having their personal information used.[105] In the United States, Google has launched a program called "Google Crossmedia Panel,"[106] which enables users to earn money in exchange for information about their web-surfing habits.

Although the personal data economy is burgeoning, it appears to be based more on contractual models than on any underlying ownership right in personal information. While there is no evidence of any ownership rights particular to this context, it is one in which heavy regulation gives individuals some degree of control, in some circumstances, to their personal information, which in turn bolsters the capacity to enter into contracts about access to and use of personal information.

Data Ownership Right

The commercial value and economic importance of data has inevitably led to calls for an ownership right in data. For the time being, these calls seem to be concentrated in the European Union, where

99 This paper has focused on Canadian law. Stacy-Ann Elvy offers a concise overview of US law, which suggests that there is no clear answer to this question in the United States and that the case law currently does not support individual ownership rights in personal data. Nevertheless, she notes that the evolving markets for consumer data and new start-ups giving consumers opportunities to commodify their data might spur a different evolution of the case law. See Stacy-Ann Elvy, "Commodifying Consumer Data in the Era of the Internet of Things" (2018) 59 BC L Rev 423 at 463ff.

100 See e.g. John Hagel III & Jeffrey F Rayport, "The Coming Battle for Customer Information", *Harvard Business Review* (1 January 1997), online: <https://hbr.org/1997/01/the-coming-battle-for-customer-information>.

101 Kenneth C Laudon, "Markets and Privacy" (1996) 39:9 Commun ACM 92–104, DOI: <10.1145/234215.234476>; Pamela Samuelson, "Privacy as Intellectual Property" (1999) 52 Stan L Rev 1125.

102 Laudon, *supra* note 101 at 104; David S Evans, "The Online Advertising Industry: Economics, Evolution, and Privacy" (2009) 23:3 J Economic Perspectives 37.

103 See e.g. Wolfgang Kerber, "Digital Markets, Data, and Privacy: Competition Law, Consumer Law, and Data Protection", *Gewerblicher Rechtsschutz und Urheberrecht Internationaler Teil (GRUR Int)* (30 April 2016) at 16, online: <https://papers.ssrn.com/abstract=2770479>; Juan Pablo Carrascal et al, "Your Browsing Behavior for a Big Mac: Economics of Personal Information Online" (Proceedings of the 22nd International Conference on World Wide Web, New York, NY, 13–17 May 2013) at 189, DOI: <10.1145/2488388.2488406>.

104 See e.g. Michael Haupt, "Introducing Personal Data Exchanges & the Personal Data Economy", *Medium.com* (7 December 2016), online: <https://medium.com/project-2030/what-is-a-personal-data-exchange-256bcd5bf447>.

105 See Stacy-Ann Elvy, "Paying for Privacy and the Personal Data Economy" (2017) 117:6 Colum L Rev 1369.

106 Google, "Google Crossmedia Panel", online: <https://crossmediapanel.com/>.

the idea has been floated and is being discussed.[107] Jeffrey Ritter and Anna Mayer argue that a copyright law framework, although evolving to protect data in some contexts, is ultimately inadequate for the task of addressing data ownership in a big data economy. They note that "these enormous data sets have nothing to do with the creative artistic assets that copyright law serves to protect. The data are industrial in nature, generated by vast networks of sensors that observe and record the smallest units of entire global supply chains."[108] The concern over the nature of the data seems overstated, as copyright law is regularly used to protect "utilitarian" works that are far from creative in nature and has proven itself to be remarkably adaptable.

Creating a data ownership right would be extremely challenging. If such a right were to be created, it would be necessary to define data for the purposes of its application. As discussed above, this would not be easy to do. Further, locating ownership may prove challenging.[109] Data are often something in which there can be multiple interests. Even in the case of personal information, it is possible to conceive of competing interests in some data. For example, it is possible to argue that a person's medical history, including their DNA, might also be the personal information of that person's children. Ownership rights seem a blunt tool to address competing interests. Problems would arise across all contexts. If there is a data ownership right, how would such a right reflect factors such as the interests of a company that collects personal information and the interests of the data subjects in their personal information collected by that company? Is the right based on the source of the information or the investment of resources in defining the parameters of and harvesting that information? In the smart cities context, how do you factor in the interests of a company that supplies the hardware that captures data, the company that derives data from the captured data, the source of any other data used in the process of deriving new data, and the city that provides access to its streets and spaces in order to collect the data? Creating a new right would require some advance consideration of such complexities. It would also require consideration and elaboration of the necessary users' rights and the need to accommodate the broad public interest in access to and use of data.[110]

Perhaps one of the fundamental problems with creating a data ownership right is the fact that there can be so many competing interests in data. These interests are present not just in the collection or creation of the data but also in its use. Copyright law has maintained the principle that facts are in the public domain largely because there is a strong and complex public interest in this being the case. There is a significant public interest in facts being free to fuel new innovation or knowledge creation. There is also a public interest in individuals being free to exchange and share facts without risk of legal constraint. Facts are an essential component of expression. While the concept of data may be more complex than facts, and may involve more human agency, data still fuel innovation, creativity, research and expression. Any new law that supported data monopolies would risk running counter to the public interest and could stifle both innovation and expression.

The discussion of the need for a broader *sui generis* data ownership right is currently concentrated in the European Union, although even in this context, it appears to be losing support. This seems comparable to the 1990s debate over the need for *sui generis* database protection. Although Europe chose to pursue this route, Canada and the United States did not. The choice not to enact *sui generis* legislation to protect databases did not have the forecasted dire effects on the development of the database industry. Given the complexities in defining data, the challenges with locating ownership and with balancing competing interests in data, as well as the need to establish significant rights of access and use, it might be preferable to allow existing law to continue to evolve through the application of established principles, IP regimes and commercial devices such as contracts.

107 See e.g. Hoeren, *supra* note 46; Bernt Hugenholz, "Data Property: Unwelcome Guest in the House of IP" (Paper presented at Trading Data in the Digital Economy: Legal Concepts and Tools, Münster, Germany, 2017), online: <https://pure.uva.nl/ws/files/16856245/Data_property_Muenster.pdf>; Thomas J Farkas, "Data Created by the Internet of Things: The New Gold Without Ownership?" (2017) 23 Revista La Propiedad Inmaterial 5.

108 Jeffrey Ritter & Anna Mayer, "Regulating Data as Property: A New Construct for Moving Forward" (2017-2018) 16 Duke L & Tech Rev 220 at 222.

109 See e.g. Farkas, *supra* note 107 at 7, 14.

110 See e.g. Scassa, "Sharing Data", *supra* note 91.

Data Sovereignty

The term data sovereignty is used in different contexts. In some cases, it is used to refer to data localization practices. Andrew Clement, for example, has linked data localization to network sovereignty in Canada. He links Canadians' loss of control over their networks to a loss of control over data; namely, "where it flows, who has access to it and what is done with it."[111]

A broader concept of control over data and its links to fundamental sovereignty issues has been the subject of considerable development in the hands of Indigenous communities in Canada, the United States, Australia and New Zealand.[112] A 2014 document[113] laid the groundwork for an approach to data sovereignty based upon the principles of ownership, control, access and possession (OCAP). The OCAP principles are holistic and address not only sovereignty over data about Indigenous peoples and communities but also sovereignty over the processes that determine what data is to be collected, by whom and how. They also address the lack of disaggregated data available about Indigenous peoples and their communities, a factor that makes it difficult to address some of the issues and challenges they face. These principles embrace ongoing stewardship of data, community capacity building and control over physical data infrastructure. Data sovereignty concerns and objectives are present in the British Columbia First Nations' Data Governance Initiative (BCFNDGI)[114] and the National Inuit Strategy on Research.[115]

Indigenous data sovereignty movements are directed at ensuring greater control over all aspects of data governance by Indigenous communities. In this sense, they are not challenges to existing rules of data ownership; rather, they are challenges to the current location of ownership of some data. Indeed, the Indigenous data sovereignty movement could have implications for policies and principles around the location of ownership and control of certain categories and types of data about Indigenous peoples. The BCFNDGI signals the potential need for legislation, data sharing and data governance agreements.[116] The National Inuit Strategy on Research calls for "processes, protocols, standards, and agreements that allow for the safe sharing of certain information and for the respectful incorporation of Inuit knowledge in data management and sharing design and implementation."[117]

Challenges with Data Ownership Rights

It is clear that the growing economic importance of data, including personal information, has given rise to increased discussion about property rights in data. From an industry perspective, property rights in data are seen to support the investment made not just in the collection of data but in its creation/generation. From a data protection perspective, a property rights basis for individual control over personal information is seen by some as a bulwark against unauthorized collection and use of personal information. Within this context, it is important to ask whether Canada's existing law provides sufficient protection for data, and, if not, what more is needed to ensure the protection of investments in data, the proper protection of personal information, and the necessary balance of ownership rights with rights of access to and use of data in the public interest.

In the 1990s, it was believed by many that without specific laws to protect property rights in databases, the database industry would founder. The choice was made in the European Union to implement such protection; in North America, by contrast, database producers were left to rely on a patchwork of laws that included copyright and contract law.

111 Andrew Clement, "Canadian Network Sovereignty: A Strategy for Twenty-First-Century National Infrastructure Building" CIGI, Data Governance in the Digital Age Special Report, 26 March 2018, online: <www.cigionline.org/articles/canadian-network-sovereignty>.

112 In New Zealand, see the Maori Data Sovereignty Network, online: <www.temanararaunga.maori.nz/>; in the United States, see the US Indigenous Data Sovereignty Network, online: <http://usindigenousdata.arizona.edu/>; and in Australia, the Maiamnayri Wingara Aboriginal and Torres Strait Islander Data Sovereignty Group is active on these issues.

113 FNIGC, OCAP, *supra* note 10.

114 BCFNDGI, online: <www.bcfndgi.com/>.

115 Inuit Tapiriit Kanatami (ITK), *National Inuit Strategy on Research* (Ottawa: ITK, 2018), online: <https://itk.ca/wp-content/uploads/2018/03/National-Inuit-Strategy-on-Research.pdf>.

116 BCFNDGI, *supra* note 112 at 8.

117 Inuit Tapiriit Kanatami, *supra* note 113 at 21.

Ultimately, it is not clear that any new law was needed. The EU experience may be instructive.

Copyright jurisprudence is slowly evolving to adopt a more complex approach to rights in data. The jurisprudence does not categorically exclude the possibility that some data are original enough to constitute protected works. Further, even if data themselves are not works, compilations of highly original data might be well protected, as they may demonstrate a strongly original selection. At the same time, the principles that have precluded the protection of facts and ideas in copyright law may continue to serve the public interest in access to and use of data that are less original to the producer and that form the building blocks of research, innovation, knowledge and expression. When users' rights are added to the mix, copyright law might offer a flexible framework well suited to the complex web of interests in data.

While copyright may offer a basis for the protection of rights in data, it has some drawbacks that are worthy of attention. Copyright law evolves relatively slowly, and while the case law shows the capacity for the protection of data, it is uneven and unpredictable. One solution is a negotiated consensus at the international level on whether, and to what extent, data should be protected under copyright law. Users' rights, while providing the framework to balance important public interests against claims of ownership, are also uncertain and unpredictable. Perhaps more importantly, many user constituencies lack the resources to fight the legal battles necessary to establish fair dealing. This creates the potential for the development of the law in a way that could unduly limit access and reuse in the broader public interest. Most battles over copyright in data to date have arisen between commercial competitors. Users' rights considerations may be marginalized or undermined in case law that evolves uniquely in this context.

Copyright is a "one-size-fits-all" regime, and it is important to recognize, as well, that strong measures adopted to combat the unauthorized reproduction and dissemination of digital works in the entertainment sector will also apply to data and compilations of data. Unduly rigid protection for TPMs, for example, will enhance the ability to assert rights over data, while at the same time precluding a more careful balancing of interests.

One alternative to allowing copyright law to develop, so as to accommodate the protection of data, is to create a data ownership right. While it might be tempting to start from scratch with a new right in data, the challenges are somewhat daunting. One primary challenge will be defining the data in which ownership rights can subsist. Another will be setting the rules for situating ownership — particularly given the complex ways in which data may be co-created. The establishment of a new regime will carry with it the risk of getting it wrong — and, as a result, of unduly burdening an industry that has thrived on fast-paced and flexible innovation. In addition to challenges of defining rights and locating ownership, maintaining an appropriate balance between ownership rights and the public interest in fostering innovation, and supporting research, criticism, free expression, education and creativity will be challenging.

Any creation of new ownership rights in data will also have implications for personal information. Data protection laws already recognize individual interests in their personal information, although these interests currently stop short of ownership rights. If ownership rights are to be recognized in data more generally, it will be difficult to exclude individuals from ownership of the personal information they generate simply by living their lives. While there is currently considerable interest in new means of managing personal information by creating personal information markets, these frameworks are still evolving outside a formal ownership regime. Once again, a new ownership right may reduce flexibility and increase complexity.

Within this rapidly evolving data environment, and with flexible and adaptable legal tools and principles already in place, a cautious "wait-and-see" approach is preferable to the creation of a new *sui generis* right. This more cautious approach does not mean that nothing can or should be done. As noted earlier, there is room to provide more structure and guidance to courts — either in domestic law or through negotiation at the international level — as to how the rules of copyright law should apply to data. Attention should also be paid to ensuring that users' rights are not neglected; the nature of data is such that any monopoly rights should be carefully limited to ensure fair rights of access and reuse in the public interest. Finally, any measures such as protection for TPMs should have their impact assessed in relation to rights in data, and any new proposed copyright reforms should also be carefully assessed in light of the needs of owners and users of data.

Author's Note

I would like to thank the two anonymous peer reviewers for their helpful comments. Many thanks as well to Bassem Awad for his feedback and suggestions.

About CIGI

We are the Centre for International Governance Innovation: an independent, non-partisan think tank with an objective and uniquely global perspective. Our research, opinions and public voice make a difference in today's world by bringing clarity and innovative thinking to global policy making. By working across disciplines and in partnership with the best peers and experts, we are the benchmark for influential research and trusted analysis.

Our research programs focus on governance of the global economy, global security and politics, and international law in collaboration with a range of strategic partners and support from the Government of Canada, the Government of Ontario, as well as founder Jim Balsillie.

À propos du CIGI

Au Centre pour l'innovation dans la gouvernance internationale (CIGI), nous formons un groupe de réflexion indépendant et non partisan doté d'un point de vue objectif et unique de portée mondiale. Nos recherches, nos avis et nos interventions publiques ont des effets réels sur le monde d'aujourd'hui car ils apportent de la clarté et une réflexion novatrice pour l'élaboration des politiques à l'échelle internationale. En raison des travaux accomplis en collaboration et en partenariat avec des pairs et des spécialistes interdisciplinaires des plus compétents, nous sommes devenus une référence grâce à l'influence de nos recherches et à la fiabilité de nos analyses.

Nos programmes de recherche ont trait à la gouvernance dans les domaines suivants : l'économie mondiale, la sécurité et les politiques mondiales, et le droit international, et nous les exécutons avec la collaboration de nombreux partenaires stratégiques et le soutien des gouvernements du Canada et de l'Ontario ainsi que du fondateur du CIGI, Jim Balsillie.

Centre for International
Governance Innovation

67 Erb Street West
Waterloo, ON, Canada N2L 6C2
www.cigionline.org

@cigionline