Data Producer's Right and the Protection of Machine-Generated Data

Peter K. Yu
Texas A&M University School of Law, peter_yu@msn.com

Follow this and additional works at: https://scholarship.law.tamu.edu/facscholar
Part of the Intellectual Property Law Commons, and the International Law Commons

Recommended Citation
Available at: https://scholarship.law.tamu.edu/facscholar/1318

This Article is brought to you for free and open access by Texas A&M Law Scholarship. It has been accepted for inclusion in Faculty Scholarship by an authorized administrator of Texas A&M Law Scholarship. For more information, please contact areteen@law.tamu.edu.
Data Producer’s Right and the Protection of Machine-Generated Data

Peter K. Yu*

I. INTRODUCTION............................................................................. 860
II. PAST LESSONS.............................................................................. 867
   A. Historical Insights.............................................................. 873
   B. Counterfactual Insights.................................................. 879
III. PRESENT NEEDS........................................................................... 884
   A. Technological Needs......................................................... 888
   B. Business Needs................................................................. 889
   C. Scientific Needs................................................................. 892
   D. Personal Needs................................................................. 893
   E. Summary............................................................................... 895
IV. FUTURE DEVELOPMENTS............................................................. 896
   A. Endogenous Complications................................................ 896
      1. Modality of Protection.................................................. 898
      2. Allocation of Rights...................................................... 901
      3. Duration of Protection................................................ 908
      4. Scope of Protection...................................................... 910
      5. Limitations and Exceptions........................................ 913
   B. Exogenous Complications................................................. 917
      1. Intellectual Property..................................................... 918
      2. Privacy........................................................................... 920
      3. Trade............................................................................. 921
      4. Investment..................................................................... 923

* © 2019 Peter K. Yu. Professor of Law, Professor of Communication, and Director, Center for Law and Intellectual Property, Texas A&M University. Earlier versions of this Article were presented at the 18th Intellectual Property Scholars Conference at the U.C. Berkeley School of Law; the 10th Annual Conference on Innovation and Communications Law at the School of Law, Bocconi University in Milan, Italy; the 7th International Intellectual Property Scholars Roundtable at Duke University School of Law; and the Workshop on “Legal Implications of the Platform Economy” at the Institute for Civil and Business Law, Vienna University of Economics and Business in Vienna, Austria. The Author would like to thank the participants of these events, in particular Frederick Abbott, Margaret Chon, Brett Frischmann, Justin Hughes, Katja Lindroos Weckström, Peter Jaszi, Nicholson Price, Jerome Reichman, and Joshua Sarnoff for valuable comments and suggestions. He is also grateful to Clemens Apf and Philipp Homar for their kind invitation to the Vienna University of Economics and Business and to Bernt Hugenholtz for providing a very timely critique that inspired this Article.
I. INTRODUCTION

Data is the new oil of today's economy—so claim policy makers, commentators, and industry analysts. In a data-driven economy, there is enormous value in the data generated and collected by humans and machines—both consciously and subconsciously. As the European Commission estimated, the value of the European data economy will more than double from €257 billion in 2014 to €643 billion by 2020. Both the McKinsey Global Institute and the Organisation for Economic Co-Operation and Development have released pioneering studies documenting the many opportunities provided by big data analytics and the emerging data-driven economy. The increase in data value,
and the attendant explosion of data,\(^5\) was the result of a confluence of factors—most notably “the rise of ‘smart manufacturing’, . . . the economic potential of ‘mining’ Big Data[,] . . . and the promise of the Internet of Things.”\(^6\)

With the arrival of big data analytics, data have also become highly valuable for uses other than what data producers or collectors initially intended.\(^7\) Because data can be used alone or in combination with previously nonexistent or inaccessible data,\(^8\) a considerable part of the value of the generated data now lies in their “option value,” not intended value.\(^9\) In an environment with ubiquitous communication, the proliferation of Internet-of-Things devices,\(^10\) and the drastic reduction

\(^{5}\) See Holmes, supra note 1, at 1-13 (discussing the data explosion); Mayer-Schönberger & Cukier, supra note 1, at 70-72 (discussing the transformation brought about by the “data deluge”); Wolfgang Kerber, *A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis*, 65 GEWERBECHRECHTINTERNATIONALER TEIL [GRUR INT] 989, 992 (2016) (Ger.) (“The amount of collected data is increasing exponentially, and it is widely expected that through the spreading of sensor technology and the ‘internet of things’ this trend will continue for the foreseeable future.”).


\(^{7}\) See Mayer-Schönberger & Cukier, supra note 1, at 153 (“[I]n a big-data age, most innovative secondary uses haven’t been imagined when the data is first collected.”); Mark Burdon & Mark Andrejevic, *Big Data in the Sensor Society*, in *BIG DATA IS NOT A MONOLITH* 61, 69 (Cassidy R. Sugimoto et al. eds., 2016) (noting that the value in data “is provided by the fact that personal data can be aggregated with that of countless other users (and things) in order to unearth unanticipated but actionable research findings”); Margaret Foster Riley, *Big Data, HIPAA, and the Common Rule: Time for Big Change?*, in *BIG DATA, HEALTH LAW, AND BIOETHICS* 251, 251 (L. Glenn Cohen et al. eds., 2018) (“The analysis of Big Data related to healthcare is often for a different purpose than the purpose for which the data were originally collected.”).

\(^{8}\) See Mayer-Schönberger & Cukier, supra note 1, at 107 (“Sometimes the dormant value can only be unleashed by combining one dataset with another, perhaps a very different one.”); Jerome H. Reichman et al., *GOVERNING DIGITALLY INTEGRATED GENETIC RESOURCES, DATA, AND LITERATURE: GLOBAL INTELLECTUAL PROPERTY STRATEGIES FOR A REDESIGNED MICROBIAL RESEARCH COMMONS* 322 (2016) (“In th[e] new research environment, scientists increasingly rely on automated knowledge discovery tools to mine and recombine vast amounts of data and literature that are flowing at rates that exceed the capacity of a single investigator to comprehend and manage.”); Burdon & Andrejevic, supra note 7, at 63 (“[E]ven seemingly irrelevant data may become relevant at some point in the future—when they can be correlated with new data sets.”).

\(^{9}\) Mayer-Schönberger & Cukier, supra note 1, at 122; see id. at 99 (“In the digital age, data shed its role of supporting transactions and often became the good itself that was traded. In a big-data world, . . . [d]ata’s value shifts from its primary use to its potential future uses.”).

\(^{10}\) As defined in the 2014 big data report from the Obama Administration:
of data storage costs, it is only logical that machines produce or collect as many data as they can. After all, it is much better to err on the side of overgeneration. As Woodrow Hartzog wrote succinctly, the ‘Internet of Things’ is a term used to describe the ability of devices to communicate with each other using embedded sensors that are linked through wired and wireless networks. These devices could include your thermostat, your car, or a pill you swallow so the doctor can monitor the health of your digestive tract. These connected devices use the Internet to transmit, compile, and analyze data.

EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 2 (2014); see also Commission Staff Working Document on the Free Flow of Data and Emerging Issues of the European Data Economy, at 41, SWD (2017) 2 final [hereinafter Commission Staff Working Document] (‘‘[Internet of Things] is a wide-ranging ecosystem of physical objects connected to the Internet, capable of identifying themselves and communicating data to other objects with the help of a communication network for digital processing.’’). See generally SAMUEL GREENGARD, THE INTERNET OF THINGS (2015) (discussing the Internet of Things). One commentator captured vividly the drastically increased production and collection of data in what commentators have referred to as the Internet of Things today:

The increase of data creation is further catalysed by new technologies and connected devices; it is often referred to as the ‘‘internet of things’’ . . . . Wearables connected to a smartphone create data to people’s physical activity. Data related to heart rates, steps taken in a certain time frame, etc., are created.[] Smart thermostats in our homes use sensors, real-time weather forecasts, and the actual activity [in] homes during the day to reduce energy usage. Networked cars have multiple sensors, steering devices and technology that can communicate with other devices outside the car. A networked car might communicate details about traffic, favourite routes and road conditions to the car owner, manufacturer, navigation service providers, insurers, construction authorities and other companies . . . . Household devices such as refrigerators may be able to order food products when scanning that we run low of our milk supply.


11. See EXEC. OFFICE OF THE PRESIDENT, supra note 10, at 4 (noting the ‘‘declining cost of collection, storage, and processing of data’’); MAYER-SCHONBERGER & CUKIER, supra note 1, at 106 (‘‘Some companies may have collected data, used it once (if at all), and just kept it around because of low storage cost—in ‘data tombs,’ as data scientists call the places where such old info resides.’’).

12. See Burdon & Andrejevic, supra note 7, at 74 (‘‘This is the business model of the contemporary commercial Internet: collect all available data because they might come in handy for new and unanticipated uses.’’). Commentators have noted that the increase in data collection is not only different in degree, but also different in kind. See GREENGARD, supra note 10, at 64 (‘‘When hundreds or thousands of sensors connect to one another, it’s possible to view data at a much higher resolution and understand relationships and patterns in a much more detailed way.’’); Scott R. Peppet, Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent, 93 TEX. L. REV. 85, 93 (2014) (‘‘Computer scientists have long known that the phenomenon of ‘sensor fusion’ dictates that the information from two disconnected sensing devices can, when combined, create greater information than that of either device in isolation.’’).

13. See Burdon & Andrejevic, supra note 7, at 62 (noting the ‘‘advent of the ‘sensor society,’ . . . in which a growing range of spaces and places, objects and devices, continuously collect data about anything and everything’’).
"In the world of big data, more is always better."\textsuperscript{14} Conscious of this increasing eagerness to produce or collect data, some commentators suggested that the Internet of Things has now slowly become the "Internet of Everything."\textsuperscript{15}

In October 2017, the European Commission proposed a new data producer's right for nonpersonal, anonymized machine-generated data.\textsuperscript{16} Driven in large part by the automotive industry and strongly supported by German commentators,\textsuperscript{17} this proposal called for the creation in data producers a "right to use and authorise the use of non-

\textsuperscript{14} Woodrow Hartzog, Privacy's Blueprint: The Battle to Control the Design of New Technologies 51 (2018); see also Mayer-Schönberger & Cukier, supra note 1, at 100 ("[I]n the age of big data, all data will be regarded as valuable, in and of itself.").

\textsuperscript{15} See Burdon & Andrejevic, supra note 7, at 72 ("The seemingly never-ending improvements in what sensors can detect, combined with the density of sensor implementation, will lead to ever-increasing forms of identification and monitoring. Sensor density, the combining of an increasing number of sensors in single objects and environments, will lead to the generation of data about everything, everywhere."); Fred H. Cate, Big Data, Consent, and the Future of Data Protection, in Big Data Is Not a Monolith, supra note 7, at 3, 3 ([D]ata are collected by sensors, which surround us in smartphones, tablets, laptops, wearable technologies and sensor-enabled clothing, RFID-equipped passports, cars, homes, and offices."); Yann Ménibre & Ilja Rudyk, The Fourth Industrial Revolution from the European Patent Office Perspective, in Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data 29, 33 (Xavier Seuba et al. eds., 2018) [hereinafter Intellectual Property and Digital Trade] ("The variety and ubiquity of the sensors embedded in connected objects make it possible to collect data of virtually any type and origin and to aggregate them into 'big data,' the raw material of [the fourth industrial revolution]."); Paul Ohm & Scott Peppet, What if Everything Reveals Everything?, in Big Data Is Not a Monolith, supra note 7, at 45, 46 ("Particularly with the rise of cheap, distributed, networked sensors—fueling the so-called Internet of Things—data sets about inanimate objects will bear the ghostly traces of the human beings who passed by or interacted with those sensors . . . ."); Peppet, supra note 12, at 89 n.14 (attributing the phrase "Internet of Everything" to Cisco CEO John Chambers).

\textsuperscript{16} Commission Communication, supra note 2, at 13. As the European Commission explained:

Machine-generated data is created without the direct intervention of a human by computer processes, applications or services, or by sensors processing information received from equipment, software or machinery, whether virtual or real.

Machine-generated data can be personal or non-personal in nature. Where machine-generated data allows the identification of a natural person, it qualifies as personal data with the consequence that all the rules on personal data apply until such data has been fully anonymised (e.g. location data of mobile applications).

\textsuperscript{17} See Hugenholtz, supra note 6, at 48 (noting that the "calls for the introduction of a novel property right in data" were in response "to demands of the automotive industry . . . and encouraged by a fair number of German lawyers and scholars" (footnote omitted)); see also Josef Drexler, Designing Competitive Markets for Industrial Data—Between Propertisation and Access, 8 J. INTELL. PROP. INFO. TECH. & ELECTRONIC COM. L. 257, 259 (2017) (noting the positions of the Bundesverband der Deutschen Industrie (German Industry Association) and the Vereinigung der Bayerischen Wirtschaft (Bavarian Industry Association)).
As the Commission explained in its Communication on "Building a European Data Economy":

This approach would aim at clarifying the legal situation and giving more choice to the data producer, by opening up the possibility for users to utilise their data and thereby contribute to unlocking machine-generated data. However, the relevant exceptions would need to be clearly specified, in particular the provision of non-exclusive access to the data by the manufacturer or by public authorities, for example for traffic management or environmental reasons. Where personal data are concerned, the individual will retain [the] right to withdraw his [or her] consent at any time after authorising the use. Personal data would need to be rendered anonymous in such a manner that the individual is not or no longer identifiable, before its further use may be authorised by the other party.19

In the run-up to this proposal and following the release of the European Commission’s communication document, commentators and consumer advocates have heavily criticized the Commission’s effort to create new sui generis protection for machine-generated data.20 One of the most vociferous critics is Professor Bernt Hugenholtz of the University of Amsterdam in the Netherlands.21 As he warned in a timely analysis:

[Introducing ... an all-encompassing property right in data would seriously compromise the system of intellectual property law that currently exists in Europe. It would also contravene fundamental freedoms enshrined in the European Convention on Human Rights and the EU Charter [of Fundamental Rights], distort freedom of competition and freedom of services in the EU, restrict scientific freedoms and generally undercut the promise of big data for European economy and society.22

Thus far, the discussion on the proposed data producer’s right has been limited to the European Union. Nevertheless, the United States has actively explored policies in response to changes in the data-driven economy and the growing significance of machine-generated data. The
2019] DATA PRODUCER’S RIGHT 865

Obama Administration, for instance, released multiple reports highlighting the complex policy questions in this economy.23

Moreover, intellectual property developments in the European Union frequently travel across the Atlantic to affect legislative and policy debates in the United States. While Congress has thus far declined to follow the European Union’s lead in adopting sui generis database protection,24 it created parity between the EU and U.S. copyright terms by extending their duration for twenty years.25 Following the October 1995 adoption of the Directive on the Protection of Individuals with Respect to the Processing of Personal Data and on the Free Movement of Such Data,26 which is commonly referred to as the Data Protection Directive, the United States also developed a “safe harbor” privacy framework.27 That framework lasted for more than a decade until the Court of Justice of the European Union rejected it as noncompliant with the Directive in October 2015.28

Given the potential for the current EU proposal for a new data producer’s right to eventually emerge on U.S. soil, this Article takes a


24. See discussion infra Part II.A.


preemptive approach of critically examining the proposal before it begins to gain traction on this side of the Atlantic. Specifically, the Article shows that the proposal has failed to learn from valuable lessons from the past, does not meet the present needs of the U.S. data-driven economy, and would likely raise considerable complications with the future development of a sound and holistic data governance regime. This Article further offers suggestions on how policy makers and commentators should develop laws and policies regarding the protection of machine-generated data.

Part II of this Article revisits the past developments concerning sui generis database protection. Arguing that these developments have provided instructive lessons on the potential adverse impacts of the proposed data producer's right, this Part revisits the critiques of sui generis database protection in the late 1990s and early 2000s. Building on the European Commission's inaugural evaluation of the EU Directive on the Legal Protection of Databases (EU Database Directive), this Part shows that sui generis database protection has thus far not provided significant benefits to EU database industries. Because that Directive has remained on the books despite not generating its intended benefits, this Article utilizes the past two decades of developments surrounding the Directive to warn about the danger of introducing hastily adopted legislation without sufficient evidence of proven needs and future success.

Part III explores whether the present technological, business, scientific, and personal needs in the United States could lend support to the proposal for a new data producer's right. When the EU Database Directive and the U.S. sui generis database protection bills were considered in the mid- to late 1990s, the Internet had only begun to enter the mainstream. Recent years, however, have seen a dramatic transformation of the digital environment, raising new legal and policy questions concerning cloud computing, big data analytics, the Internet of Things, machine learning, and artificial intelligence. This Part shows that, although data have become increasingly important and

---

29. This Part includes materials that have been updated and expanded upon from the Author's earlier discussion. See Peter K. Yu, The Political Economy of Data Protection, 84 CHI.-KENT L. REV. 777, 780-82, 789-801 (2010).
32. See Peter K. Yu, Fair Use and Its Global Paradigm Evolution, 2019 U. ILL. L. REV. 111, 121 (listing the various technological advances).
valuable, the proposal for a new data producer’s right does not meet present technological needs. Worse still, that proposal is inconsistent with the needs of our business and scientific communities. It may even cause harm to individuals and society at large.

Part IV focuses on the potential complications regarding the future development of a sound and holistic data governance regime. This Part breaks these complications down into two subgroups: endogenous complications and exogenous complications. Within the area of machine-generated data, the proposal for a new data producer’s right would raise more questions than answers. Even if this right were to generate incentives for producing these data, developing new laws to protect such data would be complicated, difficult, and, on balance, counterproductive. Outside the area of machine-generated data, complications would also arise when the protection of the proposed data producer’s right spills over into other areas of the law, including those within and outside the intellectual property system. Among the various areas impacted by the proposed right, this Part explores those concerning privacy, trade, and investment—all areas in which rights in machine-generated data could emerge. The intersection with these areas has attracted considerable attention in the intellectual property law and policy debate.

Part V concludes this Article by suggesting four courses of action that would help develop a sound and holistic data governance regime. Although the Article’s earlier analyses of the past, the present, and the future caution against the proposal for creating a new data producer’s right, this Part goes further to advance three additional suggestions beyond the outright rejection of the proposal. These suggestions focus on conceptual rethinking, international norm setting, and academic and policy engagement.

II. PAST LESSONS

When the European Commission introduced the proposal for a new data producer’s right in October 2017,33 that proposal immediately recalled the two decades of developments surrounding sui generis database protection in the European Union. Adopted in March 1996, the EU Database Directive required the then fifteen (now twenty-eight) member states to offer sui generis protection to databases that are

33. See Commission Communication, supra note 2, at 13 (advancing the proposal for the creation of a new data producer’s right for nonpersonal, anonymized machine-generated data).
created as a result of "a substantial investment in either the obtaining, verification or presentation of [database] contents." Such protection aimed to prevent the unauthorized "extraction and/or re-utilization of the whole or of a substantial part" of these contents.\footnote{Council Directive 96/9, \textit{supra} note 31, art. 7(1).}

The initial term of protection is slightly over fifteen years, "expir[ing] fifteen years from the first of January of the year following the date of completion [of the making of the database]."\footnote{Id.} Upon the demonstration of a "substantial change . . . to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations," the updated contents will constitute a new database, thereby receiving protection for at least another fifteen years.\footnote{Id. art. 10(1).} Such protection could be further extended as long as the database contents are continuously updated in a manner\footnote{Id. art. 10(3).} that would meet the "substantial new investment" threshold.\footnote{As Justin Hughes observed: One of the most disturbing things about the EU Database Directive is that the rights never seem to expire: nothing ever need enter the public domain. As long as the database producer invests in refreshing, updating, or even re-verifying the database, a new term of protection is generated over the entire database. Justin Hughes, \textit{How Extra-Copyright Protection of Databases Can Be Constitutional}, 28 U. DAYTON L. REV. 159, 211 (2003); see also REICHMAN ET AL., \textit{supra} note 8, at 339 (noting that, with the arrival of \textit{sui generis} database protection, "[p]erpetual protection . . . becomes an attainable goal for the first time in the history of intellectual property laws (disregarding, of course, trademark laws, which operate on fundamentally different principles)").}

While the proposal for the EU Database Directive garnered considerable support in the European Union in the early 1990s, leading to the Directive's eventual adoption in March 1996,\footnote{Council Directive 96/9, \textit{supra} note 31, art. 10(3).} similar proposals failed on U.S. soil.\footnote{See Mark Davison, \textit{Database Protection: Lessons from Europe, Congress, and WIPO}, 57 \textit{CASE W. RES. L. REV.} 829, 844 (2007) ("[T]he [European Community] started with a relatively minimalist model for protection and ended with a new exclusive property right that, until the [European Court of Justice's] decisions, appeared to confer considerable protection on database owners. The process by which the version that was adopted by the European Parliament in 1993 was transformed into the version that was finally adopted by the Council is somewhat opaque.").} Despite the introduction of multiple legislative bills to provide similar protection from the mid-1990s to the early

\footnote{See Yu, \textit{supra} note 29, at 780 (noting the failure of the legislative bills to introduce \textit{sui generis} database protection in the late 1990s and that the United States has yet to offer such protection).}
2000s, Congress declined to follow suit. As Charles McManis recounted:

Within ten weeks of the promulgation of the EU Database Directive, the first legislation aimed at creating a *sui generis* right in the United States, H.R. 3531, was proposed in the House of Representatives. H.R. 3531 would have created a very strong property right with a twenty-five year term and potentially severe criminal sanctions for infringement. Opposition was unexpectedly intense and the bill never reached the floor. H.R. 3531's sequel, H.R. 2652, responded to concerns from the research and library communities by including limited fair use provisions. This bill was passed by the House as part of the Digital Millennium Copyright Act [(DMCA)] on August 4, 1998; however, it was subsequently stricken during negotiations with the Senate, and the DMCA passed shortly thereafter on October 28, 1998.

The Collections of Information Antipiracy Act, H.R. 354, as introduced on January 19, 1999, would accord a database developer a broad right in factual compilations, subject not only to a specific limitation for certain non-profit educational, scientific and research uses, but also to a broader limitation for "reasonable uses" generally. The reasonableness of a particular use is to be evaluated by a multi-factor test that is somewhat reminiscent of the fair use privilege contained in section 107 of the United States Copyright Act of 1976. Notwithstanding the inclusion of this broadened fair use provision, opponents of H.R. 354 introduced an alternative bill on May 19, 1999, the Consumer and Investor Access to Information Act, H.R. 1858, that based liability on a theory more analogous to the common law tort of "hot news" misappropriation.42

Failing to obtain the much-needed support in Congress, the Clinton Administration took its proposal abroad to an international intergovernmental forum. Under the leadership of Bruce Lehman, the U.S. Patent and Trademark Office collaborated with the European Commission to push for a new Treaty on Intellectual Property in Respect of Databases (Database Treaty) that was to be created under the auspices of the World Intellectual Property Organization (WIPO).43 If adopted, this treaty would provide *sui generis* protection for at least

---


fifteen years and up to twenty-five years. The latter option, which was provided as Alternative A, would provide ten more years of protection than what the EU Database Directive currently provides. Tracking this Directive, the proposed treaty would offer new protection to any database that was modified in a manner that would result in a “substantial change to the database, ... including any substantial change resulting from the accumulation of successive additions, deletions, verifications, modifications in organization or presentation, or other alterations.”

Despite this trans-Atlantic collaboration, the treaty failed to move beyond the proposal stage. The lack of success was largely attributed to the problems posed by sui generis database protection and the continued divide between the European Union and the United States. Particularly notable was the opposition of the U.S. research and library communities, which managed to secure allied support from noted U.S. academics. The timing of the treaty negotiation was also less than

44. See World Intellectual Property Org. [WIPO], Basic Proposal for the Substantive Provisions of the Treaty on Intellectual Property in Respect of Databases to Be Considered by the Diplomatic Conference, at 31, WIPO Doc. CRNR/DC/6 (Aug. 30, 1996) (providing the draft text for art. 8(1)).

45. See id. at 30 (“The 25-year and 15-year alternatives are found in paragraph (1) and paragraph (2) of Article 8. The decision on the term of protection has been left to the Diplomatic Conference.”).

46. See Council Directive 96/9, supra note 31, art. 10 (providing sui generis database protection for at least fifteen years).

47. WIPO, supra note 44, at 31.

48. As Pamela Samuelson recounted:
Widespread criticism of the Chairman’s draft database treaty caused its quick removal from the conference agenda. Consideration of the draft database treaty has, however, not merely been postponed or sent back to the Committee of Experts for further refinement; it has been taken off the table. In order for database protection issues to be raised again at WIPO, the governing body of WIPO will have to constitute a new Committee of Experts to study the matter.

Samuelson, supra note 43, at 426 (footnote omitted); see also Jonathan Band & Brandon Butler, Overlapping Forms of Protection for Databases, in OVERLAPPING INTELLECTUAL PROPERTY RIGHTS 189, 206 (Neil Wilkof & Shamad Basheer eds., 2012) (“In the face of the opposition from the developing countries and the United States, the WIPO governing body decided at the outset of the Diplomatic Conference to defer further consideration of the database treaty.”).

49. As Jonathan Band and Brandon Butler recounted:
The scientific community in the United States reacted with alarm to the proposed [Database Treaty], arguing that it would stifle research. The science agencies within the US government (e.g., the Environmental Protection Agency and the National Oceanographic and Atmospheric Administration) took the lead in persuading the National Economic Council within the White House to oppose adoption of a database treaty.
ideal: the Database Treaty proposal was cramped into an already crowded negotiation agenda that included two important Internet treaties—the WIPO Copyright Treaty and the WIPO Performances and Phonograms Treaty.

Had Commissioner Lehman and his staff managed to get the proposed Database Treaty adopted at WIPO, they would have achieved what commentators have referred to as “policy laundering”—the “efforts [undertaken] by policy actors to have policy initiatives seen as exogenously determined, or even seen as requirements imposed by powerful others.” By negotiating an international agreement to cover protection that did not receive sufficient domestic legislative support, the Clinton Administration sought to launder an unpopular and ill-advised policy through the international negotiation process. In doing so, the Administration hoped that the policy would become more legitimate when it returned to the home soil for deliberation. Such a laundering effort aimed to make an end run around Congress and the domestic deliberative process.

Subpart A begins with sui generis database protection because such protection has attracted similar criticisms as the proposal for a new data producer’s right. Putting the debate on data protection in a

---

Thus, Commissioner Lehman arrived in Geneva in early December 1996 with instructions to stop the very database treaty that he had helped set in motion. Band & Butler, supra note 48, at 206; see also Samuelson, supra note 43, at 423-25 (discussing the expressions of concern and opposition to the draft Database Treaty from various quarters).

50. See generally Samuelson, supra note 43 (discussing the 1996 WIPO Diplomatic Conference in Geneva and the formation of the WIPO Internet Treaties).


55. See Hosein, supra note 53, at 188 (noting the general notion that “international cooperation is inherently good” and the general belief that international cooperation is “seen as benign and ... for the most part uninterrogated”).

56. As Samuelson reminded us:

Had [the development of the Database Treaty] succeeded in Geneva, Clinton administration officials would almost certainly have then argued to Congress that ratification of the treaties was necessary to confirm U.S. leadership in the world intellectual property community and to promote the interests of U.S. copyright industries in the world market for information products and services.

Samuelson, supra note 43, at 374.
historical perspective, this subpart revisits the U.S. legislative debate concerning *sui generis* database protection in the late 1990s and early 2000s. While it is helpful to recount the different reasons why Congress rejected such protection, this historic debate also provides instructive lessons on the potential adverse impacts of the proposed data producer’s right. To a large extent, the arguments against introducing *sui generis* database protection in the United States two decades ago will militate against the present proposal for a new data producer’s right.

Because Congress declined to introduce *sui generis* database protection in the late 1990s and early 2000s, we will never know what would have happened had such protection been introduced in the first place. Nevertheless, counterfactual reasoning can be deployed to shed light on this potential scenario.57 Even better, the past two decades of developments surrounding the EU Database Directive have provided a comparable experience to support a counterfactual narrative. To document these developments, subpart B recapitulates the widely criticized evaluation the European Commission undertook shortly before the Directive’s tenth anniversary.58 Given that the Commission recommended the retention of the Directive despite a highly negative evaluation, the discussion in this subpart further illustrates the problems that would have surfaced had *sui generis* database protection been adopted in the United States. The identified problems also provide an important warning of the danger of hastily adopting legislation without sufficient evidence of proven needs and future success.


58. COMM’N OF THE EUROPEAN COMMS., supra note 30.
A. Historical Insights

Following the adoption of the EU Database Directive in March 1996, Congress considered a number of *sui generis* database protection bills. The first bill—the Database Investment and Intellectual Property Antipiracy Act of 1996—was introduced less than three months after the Directive’s adoption. Since then, Congress has considered several legislative bills—most notably the Collections of Information Antipiracy Act of 1998, the Collections of Information Antipiracy Act of 1999, and the Consumer and Investor Access to Information Act of 1999. Up until the early 2000s, the U.S. policy makers and legislators remained interested in introducing *sui generis* database protection. With the passage of time, however, such efforts slowly faded away.

As the proponents of these legislative bills argued, *sui generis* protection would provide to U.S. database industries the much-needed incentives to collect data. Such protection was particularly attractive following the United States Supreme Court’s decision to drastically reduce copyright protection for databases in *Feist Publications, Inc. v. Rural Telephone Service Co.* Decided in March 1991, the Court held

60. McManis, supra note 42, at 35; see also Marci A. Hamilton, *A Response to Professor Benkler*, 15 BERKELEY TECH. L.J. 605, 615 (2000) (“In May 1996, the first database protection bill in the United States was introduced in Congress.”).
61. H.R. 2652, 105th Cong. (1998). Upon adoption by the House of Representatives, this bill became part V of the DMCA until it was dropped from the larger H.R. 2281 bill. As Jerome Reichman and Paul Uhlir recounted:

The House of Representatives adopted H.R. 2652, and the Subcommittee on Courts and Intellectual Property then attached it to the Digital Millennium Copyright Act, which became H.R. 2281, as sent to the Senate. The database portion was dropped prior to Congressional enactment of that bill, however, and it was reintroduced with some modifications as H.R. 354 in January 1999.


Database publishers argued that the Supreme Court’s decision in *Feist* significantly diminished publishers’ incentive to invest in the compilation of information. They argued that post-*Feist*, copyright was particularly ineffective with respect to large comprehensive online databases that are used by means of a search engine. The compiler has exercised no selection because the databases are comprehensive.
that the white pages of a telephone directory did not constitute a sufficiently original work of authorship that qualified for copyright protection. As Justice Sandra Day O’Connor clearly stated, a compilation did not warrant copyright protection unless the collected information was selected, coordinated, or arranged in an original manner.

Moreover, the EU Database Directive specifically includes a reciprocity provision that excludes protection from foreign database producers whose country of origin does not offer comparable protection. Thus, if U.S. database industries were to compete effectively against their EU counterparts, the United States would have to offer parity protection. Without such protection, investment for database

Further, arrangement only occurs when the user conducts a search. In the absence of selection and arrangement, copyright protection is not available.

Band & Butler, supra note 48, at 201. Justin Hughes, however, questioned the impact of Feist on the U.S. database industry:

In North America, there is no evidence that database production has weakened in the post-Feist environment and, in fact, one of the major proponents of database protection announced plans in 2000 to sell $2.5 billion worth of copyright-protected newspapers and put a substantial chunk of those proceeds into “expanding its electronic databases” in a program to make 80% of its revenue “come from the electronic distribution of information” within five years. That does not sound like a business environment inhospitable to investment in databases.


66. See Feist, 499 U.S. at 364 (“Rural’s white pages lack the requisite originality, Feist’s use of the listings cannot constitute infringement.”).

67. See id. at 358 (“[W]e conclude that the [1976 Copyright Act] envisions that there will be some fact-based works in which the selection, coordination, and arrangement are not sufficiently original to trigger copyright protection.”).

68. See Council Directive 96/9, supra note 31, art. 11(3) (“Agreements extending the right provided for in Article 7 to databases made in third countries and falling outside the provisions of paragraphs 1 and 2 shall be concluded by the Council acting on a proposal from the Commission. The term of any protection extended to databases by virtue of that procedure shall not exceed that available pursuant to Article 10.”); see also Reichman & Samuelson, supra note 43, at 96-97 (discussing the European Commission’s “strict criterion of material reciprocity”); Peter K. Yu, Currents and Crosscurrents in the International Intellectual Property Regime, 38 LOY. L.A. L. REV. 323, 379 (2004) (discussing the reciprocal provisions in the EU Database Directive).

69. As Band and Butler recounted:
production would slowly migrate from the United States to Europe.⁷⁰ Policy makers and industries also feared that "databases produced by U.S. companies [would] become vulnerable to foreign competition and piracy in European markets."⁷¹ In short, EU database producers would become more successful than those in the United States.

Notwithstanding these arguments and the related lobbying effort, Congress declined to introduce *sui generis* database protection for five reasons. First, such protection did not fit well with the U.S. copyright regime. As noted earlier, the *Feist* Court declined to grant copyright protection to nonoriginal, noncreative databases.⁷² Because of this seminal decision, and the fact that few commentators had considered the case wrongly decided, the proponents of *sui generis* database protection bills were fighting an uphill battle.⁷³ After all, any new

---

The proponents . . . argued that adoption of the Database Directive necessitated enactments of database legislation in the United States. The Database Directive’s *sui generis* protection is available only on a reciprocity basis. This meant that a non-EU publisher can receive the heightened level of protection only if the publisher’s country of origin afforded an equivalent level of protection. In other words, if the US did not enact database legislation on par with the Database Directive, then US publishers could not receive this added protection in Europe. European publishers, in contrast, would receive the protection against US publishers, thereby placing US publishers at a competitive disadvantage.

Band & Butler, *supra* note 48, at 201; *see also* Nat’l Research Council, *A Question of Balance: Private Rights and the Public Interest in Scientific and Technical Databases* 52 (1999) (noting the “perceived increased vulnerability of databases to . . . a new European legal regime that has been alleged to place U.S. database rights holders at a competitive disadvantage in Europe”).

⁷⁰ As I noted in an earlier article:

> In light of the very different protection offered in these countries, businesses have engaged in regulatory arbitrage by relocating their operations to jurisdictions that offer more favorable legal environments. To attract foreign investment and to retain local businesses, countries now actively participate in a “race” to either the top or the bottom.


⁷¹ Yu, *supra* note 27, at 623; *see also* Reichman & Samuelson, *supra* note 43, at 96 (“Databases made in countries having no similar legislation would remain vulnerable to wholesale copying within the European Union itself.”).

⁷² *Feist*, 499 U.S. at 364.

⁷³ *See* Yu, *supra* note 29, at 780-81 (“The cold reception of database protection can be largely attributed to the 1991 case of *Feist Publications, Inc. v. Rural Telephone Service* . . . .”)
protection created through these bills would result in conflicts with the then-existing copyright regime, which denied protection to databases that do not involve selection, coordination, or arrangement of information in an original manner.

Second, questions arose over the constitutionality of granting property rights in nonoriginal, noncreative databases. By offering protection under the “sweat of the brow” doctrine expressly rejected by the Feist Court, the proposed protection would unlikely pass muster under the Copyright Clause of the Constitution. The Feist Court not only rejected the protection for the white pages of a telephone directory but also explicitly stated that “[o]riginality is a constitutional requirement.” Apart from issues relating to the Copyright Clause, many commentators also subscribed to the view that the bills would raise serious constitutional questions under both the Commerce Clause and the First Amendment.
Third, many policy makers and commentators considered *sui generis* database protection to be bad public policy whose costs would likely outweigh its benefits. For example, the protection "would confer far broader and stronger exclusive rights on database content than is necessary to provide the needed incentives for database producers." By granting a monopoly over collected data, such protection would also allow private actors to lock up information that was essential to basic scientific research and future creative endeavors. The protection would even establish an anticompetitive environment that would make it difficult for value-adding products and services to enter the market. Such an environment, in turn, would make information products more expensive, thereby harming consumers and society at large.

Fourth, *sui generis* database protection was considered unnecessary because database producers already enjoyed significant protection under private contracts, the misappropriation doctrine, and state unfair competition laws. These producers could also protect their valuable data by deploying technological protection measures.

---


80. See Reichman & Samuelson, *supra* note 43, at 113-24 (discussing the adverse impact of *sui generis* database protection on scientific research and education); Reichman & Uhlir, *supra* note 61, at 796-821 (discussing the adverse impact of database protection laws on scientific, technical, and educational users of factual data and information).


82. See Jane C. Ginsburg, *Copyright, Common Law, and Sui Generis Protection of Databases in the United States and Abroad*, 66 U. CHIC. L. REV. 151, 151 (1997) ("If copyright is unavailing, contract is appearing more promising, as mass-market, 'shrinkwrap' and 'click-on' licenses gain acceptance. Indeed, if contractual protection of compiled information persists unpreempted by federal copyright law or policy, it may provide more effective protection than did copyright." (footnote omitted)); Marshall Leaffer, *Database Protection in the United States Is Alive and Well: Comments on Davison*, 57 CASE W. RES. L. REV. 855, 858-59 (2007) (discussing the alternative form of protection provided by state law remedies, in particular contract law remedies).

83. See Ginsburg, *supra* note 82, at 157-64 (discussing the protection provided by the common law doctrine of misappropriation to databases).


85. See NAT'L RESEARCH COUNCIL, *supra* note 69, at 64-68 (discussing the availability of technological protections, including encryption, watermarks, online access controls, and trusted systems); Band & Butler, *supra* note 48, at 200 ("The Digital Millennium Copyright Act provides additional protection to the publishers of electronic databases. For databases
Although Congress had yet to offer anti-circumvention protection through the DMCA when *sui generis* database protection bills began to emerge in the mid-1990s, technological measures had been used to protect computer software and digital content for decades before the statute’s adoption.

Finally, the proposed database protection legislation did not get sufficient traction on Capitol Hill. Out of the three major stakeholders in the global database industry, only McGraw-Hill was American. To be sure, the United States emulated the European Union when the latter extended the term of copyright protection from the life of the author plus fifty years to the life of the author plus seventy years. Nevertheless, those in the United States who would have benefited from *sui generis* database protection did not compare favorably with distributed in digital form, technological measures that prevent unauthorized access, reproduction, and distribution are becoming more prevalent and powerful.”; Drexl, supra note 17, at 291 (“Data holders are able to charge a price for making data available to third parties based on factual control over data, supported by technical protection measures.”); Leaffer, supra note 82, at 859 (“[T]he anticircumvention provisions of the [DMCA] create substantial protection, particularly for electronic databases.” (footnote omitted)); Pollack, supra note 65, at 99-111 (discussing the alternative protection provided by technological protection measures and the anti-circumvention provision of the DMCA).


87. See Peter K. Yu, Anticircumvention and Anti-Anticircumvention, 84 DENV. U. L. REV. 13, 14 (2006) (“Digital rights management . . . systems, including technological measures that are used to protect copyrighted works, are not new. They have existed for at least the last couple of decades.”) (footnote omitted).

88. As I noted in an earlier article:

Instead of providing substantial facts on the harm that the lack of Congressional responses would cause to the U.S. database industry, the sponsors and proponents of this legislation were able to make only generalized claims of potential foreign competition and piracy in European markets. Because many businesses were both producers and users of data, they remained reluctant to support stronger database protection until they were certain that the proposed legislation would strike the appropriate balance between the production of databases and the use of collected information.

Yu, supra note 29, at 782 (footnote omitted).

89. See Benkler, supra note 81, at 592 (“[T]he rallying cry ‘protect our database producers from the Europeans’ rings hollow when testimony suggested that of the three major industry stakeholders only one, McGraw Hill, is an American company, while another, Reed-Elsevier, is a European company. (The third was Thompson, a Canadian company.)” (footnotes omitted)). Ironically, as the National Research Council observed in its report, “The only significant economic analysis done in the United States with regard to the [*sui generis* database protection] legislation was an article commissioned by [these two foreign stakeholders].” NAT’L RESEARCH COUNCIL, supra note 69, at 93.

those who benefited from the copyright term extension.\textsuperscript{91} Even so, such extension had been so controversial that its constitutionality was challenged before the Supreme Court in \textit{Eldred v. Ashcroft}.\textsuperscript{92}

Taken together, these five reasons explain why the United States—or, for that matter, many other non-EU countries\textsuperscript{93}—have thus far declined to introduce \textit{sui generis} database protection. More importantly, those arguments that helped defeat database protection bills two decades ago could apply to any proposal for a new data producer’s right. Thus, if such a proposal is to emerge in the United States, policy makers should actively draw on prior lessons regarding \textit{sui generis} database protection to highlight the proposal’s weaknesses and the challenges and potential constitutional questions it will bring about.

\section*{B. Counterfactual Insights}

Although the past experiences concerning the introduction of \textit{sui generis} database protection in the United States are highly instructive, it will also be useful to look at what would have happened had such protection been put in place on U.S. soil. Because Congress declined to introduce \textit{sui generis} database protection in the late 1990s and early 2000s, counterfactual reasoning will be needed to imagine the scenario where the United States has introduced such protection. Fortunately, the EU Database Directive was adopted in March 1996 and has since been implemented by all twenty-eight EU members.\textsuperscript{94} Once cross-Atlantic differences have been taken into account, the past experience with the Directive provides useful insights into whether the United

\begin{itemize}
\item \textsuperscript{91} See Hugh Hansen et al., \textit{Panel I: Database Protection}, 11 Fordham Intell. Prop., Media \& Ent. L.J. 275, 299 (2001) ("Whereas the copyright term affected a lot more people, database is more discrete. I think you also have a user community that is much more active with regard to database protection than there was with regard to the Term Directive." (remarks of Hugh Hansen, Professor of Law, Fordham University School of Law)).
\item \textsuperscript{92} 537 U.S. 186.
\item \textsuperscript{93} See Reichman et al., supra note 8, at 325 (noting that database protection laws are "now enacted in fifty-five countries"); Band \& Butler, supra note 48, at 196 (listing Iceland, Mexico, South Korea, and Turkey as some of the countries that have adopted some form of \textit{sui generis} database protection).
\item \textsuperscript{94} As Band and Butler recounted:

Greece, Ireland, Luxembourg, and Portugal delayed in their implementation of [the EU Database Directive]. On 30 July 1999, the European Commission initiated legal proceedings before the European Court of Justice against these four countries for failure to implement the Directive by the implementation date. All 27 [now twenty-eight] member states of the European Union have . . . adopted the directive.

Band \& Butler, supra note 48, at 192.
\end{itemize}
States would have been better or worse off had the country introduced \textit{sui generis} database protection.

In February 2006, shortly before the tenth anniversary of the EU Database Directive, the European Commission undertook a comprehensive evaluation of the Directive.\textsuperscript{95} This inaugural report specifically assessed whether the Directive’s policy goals had been achieved and whether the Directive had an adverse impact on competition.\textsuperscript{96} The report’s findings were both revealing and highly disturbing.

As the report showed, the EU Database Directive failed to provide much benefit to the European Union.\textsuperscript{97} Worse still, the Directive might have harmed the EU publishing and database industries. In 2001, the year after most of the first fifteen EU member states had transposed the Directive into national law, there were 4085 EU database entries.\textsuperscript{98} By contrast, in 2004, that number had quickly declined by close to a quarter to 3095.\textsuperscript{99} Although the Directive aimed to create a level playing field between the EU and U.S. database industries, “[b]etween 2002 and 2004, the European share decreased from 33\% to 24\% while the US share increased from 62\% to 72\%. The ratio of European/US database production, which was nearly 1:2 in 1996, has become 1:3 in 2004.”\textsuperscript{100}

Given the EU database producers’ reduced competitiveness, one would naturally expect the European Commission to recommend the

\footnotesize{
\begin{itemize}
\item \textsuperscript{95} \textit{COMM’N OF THE EUROPEAN COMMS., supra} note 30; see also James Boyle, Opinion, \textit{James Boyle: Two Database Cheers for the EU}, \textit{FIN. TIMES} (Jan. 2, 2006), https://www.ft.com/content/99610a50-7bb2-1lda-ab8e-0000779e2340 (discussing the European Commission’s evaluation report).
\item \textsuperscript{96} \textit{See COMM’N OF THE EUROPEAN COMMS., supra} note 30, at 3 (“The purpose of this evaluation is to assess whether the policy goals of Directive 96/9/EC on the legal protection of databases ... have been achieved and, in particular, whether the creation of a special ‘sui generis’ right has had adverse effects on competition.” (footnote omitted)).
\item \textsuperscript{97} \textit{See id.} at 5 (“The economic impact of the ‘sui generis’ right on database production is unproven. Introduced to stimulate the production of databases in Europe, the new instrument has had no proven impact on the production of databases.”).
\item \textsuperscript{98} \textit{Id.} at 24.
\item \textsuperscript{99} \textit{Id.} Nevertheless, Matthias Leistner noted that the Commission’s evaluation completely ignore[d] the fact that in the first years after the enactment of the Database Directive the European database industry had indeed increased substantially and that the crucial decrease of database production from 2001 to 2004 might indeed rather be linked to the general crisis of the information industries in that period or other incidental factors.
\item \textsuperscript{100} \textit{COMM’N OF THE EUROPEAN COMMS., supra} note 30, at 22.
\end{itemize}
}
repeal, withdrawal, or at least amendment of the Directive. The Commission, however, declined to make any of those recommendations. Instead, it advanced three justifications for retaining this ineffective and highly problematic Directive.

First, the Commission “ha[d] received strong representations from the European publishing industry that ‘sui generis’ protection [was] crucial to the continued success of their activities.” As the Commission explicitly acknowledged, “[T]he attachment to the new right [was] a political reality that seem[ed] very true for Europe.” While the Commission’s observation was undoubtedly supported by the online surveys of stakeholders in the database industry, the survey approach was methodologically flawed. As James Boyle lamented in the Financial Times immediately after the report’s release, the Commission’s “questionnaire to the European database industry asking if they liked their intellectual property right [was comparable to] a procedure with all the rigour of setting farm policy by asking French farmers how they feel about agricultural subsidies.”

Second, a repeal of the EU Database Directive “would require withdrawing, or ‘reverse’, legislation and that might reopen the original debate on the appropriate standard of ‘originality.’” Similarly, a reformulation of the scope of the sui generis database right would “require the Community legislator to revisit the compromise underlying the two-tier protection introduced by the Directive where a distinction is made between ‘original’ databases that have to comply with a high standard of ‘originality’ and ‘non-original’ databases that enjoy a form of ‘sui generis’ protection.” The Commission’s position is understandable, considering that a key achievement of the Directive was to facilitate harmonization within the European Union. Before

---

101. See id. at 24-27 (outlining the four options as (1) “[r]epeal the whole directive,” (2) “[w]ithdraw the ‘sui generis’ right,” (3) “[a] mend the ‘sui generis’ provisions,” and (4) “[m] aintain[] the status quo”).
102. Id. at 24.
103. Id. at 25.
104. Boyle, supra note 95; see also Davison, supra note 40, at 843 (“Of course, one should take with more than a grain of salt the suggestion from an industry that gets legal protection that the continuation of the protection is critical for its survival.”).
105. COMM’N OF THE EUROPEAN COMMS., supra note 30, at 6.
106. Id.
107. As the European Commission explained in its evaluation report:

The Directive attempts to establish a uniform threshold of “originality” for “original” databases. This level of protection has the effect that the United Kingdom and Ireland, which applied a lower threshold of “originality”, were required to “lift the bar” and accord copyright protection to only those databases which were
the Directive’s full implementation, some EU countries—most notably Ireland and the United Kingdom—took copyright protection to “‘non-original’ databases involving considerable skill, labour or judgment in gathering together and/or checking a compilation.” Nordic countries also had a “catalogue rule,” which afforded protection for a short duration to “[c]atalogues, tables, and similar compilations in which a large number of particulars have been summarized.” By contrast, after the implementation of the Directive, nonoriginal databases receive only sui generis database protection, not copyright protection.

Third, “[r]emoving the ‘sui generis’ right and thereby allowing Member States to revert to prior forms of legal protection for all forms of ‘non-original’ databases that do not meet the threshold of ‘originality’, might be more costly than keeping it in place.” As the outcome of the European Commission’s evaluation revealed, “laws can be politically entrenched, and amending these laws can be difficult even if they have proven to be ineffective or harmful.” Thus, if

---

“original” in the sense of the author’s own intellectual creation. As a result, databases which qualified for copyright protection under the “sweat of the brow” regime would no longer be protected. In exchange, and in order to compensate for the loss of the “sweat of the brow” protection, the “sui generis” form of protection for “non-original” databases was introduced as an entirely novel form of intellectual property.

Id. at 8; see also William F. Patry, A Few Observations About the State of Copyright Law, in COPYRIGHT LAW IN AN AGE OF LIMITATIONS AND EXCEPTIONS 85, 100 (Ruth L. Okediji ed., 2017) (describing as “laudable” the Database Directive’s “goal . . . to harmonize the level of originality required for a database to be protected under copyright laws”). But see Davison, supra note 40, at 822 (“The lack of harmony is reflected in conflicting case law decisions and different approaches to the adoption of defenses allowed under the Directive. While the difficulties of conflicting case decisions may be partially resolved by the [European Court of Justice] decisions, in the interim, there have been considerable differences in approach both within and between individual European jurisdictions.”).

108. See COMM’N OF THE EUROPEAN COMMS., supra note 30, at 8 (noting that “the United Kingdom and Ireland . . . applied a lower threshold of ‘originality’”).
109. Id. at 7.
110. See id. at 8 n.9 (“Denmark, Finland and Sweden protected ‘a catalogue, a table or another similar production in which a large number of information items have been compiled’ under the so-called ‘catalogue rule’.”). See generally Gunnar W.G. Karnell, The Nordic Catalogue Rule, in PROTECTING WORKS OF FACT 67 (Egbert J. Dommering & P. Bernt Hugenholtz eds., 1991) (discussing the Nordic catalogue rule).
111. Karnell, supra note 110, at 67 (quoting 5 ch. 49 § LAG OM UPPHOVSRÄTT TILL LITTERÄRA OCH KONSTNÄRLIGA VERK (Svensk författningssamling [SFS] 1960:729) (Swed.).
112. COMM’N OF THE EUROPEAN COMMS., supra note 30, at 27.
113. Yu, supra note 29, at 801; see also Clayton P. Gillette, Lock-in Effects in Law and Norms, 78 B.U. L. REV. 813 (1998) (discussing the lock-in effects in law); Yu, supra note 87,
policy makers are unsure about the potential benefits or effectiveness of a new legislative proposal, such as the one to create a new data producer’s right, they should advance with great caution, demand empirical evidence, and undertake impact assessment.

In sum, the European Commission’s evaluation of the EU Database Directive provided an important warning about the danger of hastily adopting legislation without sufficient evidence of proven needs and future success. While laws that fail to achieve the intended goals or that ill fit the economic and technological environment should be repealed or modified, such changes do not always take place even with careful documentation of the laws’ problems. Indeed, as then-Dean (now Judge) Guido Calabresi rightly observed, the legal system has been filled with a “retentionist bias” that allows outmoded laws to stay on the books.

Thus, contrary to the claims advanced by the proponents of sui generis database protection, the United States’ failure to adopt such protection has actually helped local database producers. Mark at 57-61 (discussing entrenched laws and their lock-in effects, with illustrations from the anti-circumvention provision of the DMCA).

114. See Hugenholtz, supra note 6, at 70 (“Introducing a novel right of intellectual property should of course never be done in the spur of the moment.”).

115. See William Patry, How to Fix Copyright 52 (2011) (noting the need for “mandatory, independently-produced, impartial, empirically rigorous impact statements before any new copyright legislation is passed”); Hugenholtz, supra note 6, at 70 (“Any new [intellectual property] right should be contemplated only after conducting thorough economic, evidence-based research that demonstrates a real need for the right and predicts its consequences for information markets and society at large.”); Peter K. Yu, Digital Copyright and Confuzzling Rhetoric, 13 Vand. J. Ent. & Tech. L. 881, 918-22 (2011) (noting the need for the proponents of intellectual property reform to provide credible empirical support).

116. Impact assessments are increasingly important in the intellectual property field. As I noted in an earlier article:

In recent years, impact assessments have been widely endorsed in the areas of human rights, public health, and biological diversity. Assessment, evaluation, and impact studies also constitute one of the six clusters of recommendations adopted as part of the WIPO Development Agenda in October 2007. To ensure more accurate assessments, countries should deploy holistic impact assessments that involve institutional cooperation across sectors and agencies. Preferably, these assessments will be conducted before the introduction of new forms of protection. If such assessments cannot be undertaken at that time—for example, as a result of heavy external pressure from developed country governments—assessments should still be conducted following the introduction of new standards or measures, perhaps after a specified period of time.


117. See Guido Calabresi, A Common Law for the Age of Statutes 60 (1982).

118. As I observed in an earlier article:
Davison went even further to suggest that the country’s refusal to follow the European Union’s lead has benefited the world at large:

If the U.S. domino had fallen, the world would today have far more widespread database rights for non-original databases via either multilateral or bilateral agreements. And then, after it was too late, the world would have discovered what the [European Community] has now discovered, namely, that there is no need for such a right. 119

III. PRESENT NEEDS

When the debate about *sui generis* database protection emerged in the early 1990s, the technological environment was very different from what it is today. Indeed, the proposal for the EU Database Directive drew on the catalogue rule found commonly in Nordic countries. 120 Although the Directive ended up covering both electronic and nonelectronic databases, 121 the catalogue rule, as its name suggests, was designed primarily for the offline world, not the online world. 122 It


120. As Reichman and Samuelson recounted: [The Nordic countries had already experimented with short-term, copyright-like protection of noncopyrightable compilations—known as the Nordic “catalogue rule”—with a view to curbing commercial piracy without extending full copyright protection to borderline literary productions that lacked creative authorship. Accordingly, in 1992, the Commission proposed an innovative directive to protect such databases that was “loosely modeled on the Nordic catalogue rule, [and] more directly and strongly protects electronic information tools.” A greatly amended version of this proposal was adopted by the Council of Ministers and the European Parliament in July 1995 which, with only modest, technical changes, took effect on March 11, 1996. Yu, *supra* note 29, at 800-01.

121. See Council Directive 96/9, *supra* note 31, recital 14 (noting that “protection under this Directive should be extended to cover non-electronic databases”); see also COMM’N OF THE EUROPEAN COMMS., *supra* note 30, at 10 (“The proposal for the Directive was originally limited to electronic databases but now includes analogue, including hard copy or traditional print media, and electronic forms, including digital or online.”).

122. Undoubtedly, the widespread use of the Internet has a significant impact on the database industry and the figures regarding data entries. As the European Commission’s evaluation report noted:
is therefore no surprise that the European Commission’s evaluation of the EU Database Directive focused on the metrics provided by the *Gale Directory of Databases*. The Commission also sent its restricted online surveys to traditional database producers—namely, “500 European companies and organisations involved in the database industry (publishers, suppliers of data and information, database manufacturers, distributors, etc.).”

Today, however, the technological environment is dominated by not only the Internet but also cloud computing, big data analytics, the Internet of Things, machine learning, artificial intelligence, and other technological advances. One therefore cannot help but wonder whether the advent of these new technologies warrants a fresh law and policy debate in the area of data production. After all, the proposal for a data producer’s right targets the *production* of machine-generated data. Unlike the *sui generis* database right, the proposed right does not aim to provide incentives for “the obtaining, verification or presentation of [database] contents.”

In recent years, policy makers and commentators have underscored the highly important and ever-growing role of data and data-based innovation in our economy. From online platforms to

---

With respect to the overall decline of database “entries” as of 2001, the [European Association of Directory and Database Publishers (EADP)] argues that database “entries” decreased due to a shift toward the online provision of information.

The EADP further points out that database delivery has shifted from stand-alone database products, such as CD-ROMs and dedicated on-line access to specific databases, to “portal” based applications which enable a single point of access to many databases. According to the EADP, this trend is not reflected in the *Gale Directory of Databases* statistics.


123. *See id.* at 5 (“The evaluation was conducted on the basis of ... information received from the *Gale Directory of Databases* ..., the largest existing database directory which contains statistics indicating the growth of the global database industry since the 1970s.”).

124. *Id.* at 5 n.5.

125. *See Yu, supra* note 32, at 121 (listing the various technological advances).


127. As the European Commission declared in the opening paragraph of its *Communication on “Building a European Data Economy”*:

Data has become an essential resource for economic growth, job creation and societal progress. Data analysis facilitates the optimisation of processes and decisions, innovation and the prediction of future events. This global trend holds enormous potential in various fields, ranging from health, environment, food
interconnected devices, a large quantity and variety of data are being produced and collected every day. While some of these data are collected the old-fashioned way, or through automation, others require the development of, and investment in, new enabling technologies, such as sensors, processors, embedded software, data storage systems, and automated services. Indeed, without these new technologies, some of the data generated today may not be produced or collected in the first place.

Thus, it is understandable why policy makers would seriously consider offering protection to machine-generated data. The stakes of not offering adequate protection in a data-driven society are just too high. Today, data can be used and reused by many parties in all sectors. They can also be used alone or in combination with previously nonexistent or inaccessible data. In addition, data can be included in licensing arrangements or monetization deals. In short,

security, climate and resource efficiency to energy, intelligent transport systems and smart cities.

Commission Communication, supra note 2, at 2.

128. See HOLMES, supra note 1, at 16-18 (discussing the volume and variety of big data); Drexl, supra note 17, at 264 (discussing the “three Vs” of big data: “volume, velocity and variety”).

129. See Michael Bailey, Will Big Data Diminish the Role of Humans in Decision Making?, in BIG DATA IS NOT A MONOLITH, supra note 7, at 163, 163 (“The data revolution is creating new industries and revolutionizing existing ones, opening up completely new products and operational capabilities.”); Keith E. Maskus, Fostering Innovation in Digital Trade, in INTELLECTUAL PROPERTY AND DIGITAL TRADE, supra note 15, at 19, 21 (“Beyond basic e-commerce, the world is just beginning to understand the breadth and depth of digital services that will be traded across borders to facilitate growing new industries, such as autonomous driving vehicles, smart metering technologies, wearable smart fabrics, and the Internet of Things.”).

130. See Ménière & Rudyk, supra note 15, at 32 (“By 2020 it is estimated that 26-30 billion devices in the home and workplace will be equipped with sensors, processors, and embedded software, all connected to the Internet of Things.”).

131. See MANYIKA ET AL., supra note 4, at 4 (“Big data has now reached every sector in the global economy. Like other essential factors of production such as hard assets and human capital, much of modern economic activity simply couldn’t take place without it.”); ORG. FOR ECON. CO-OPERATION & DEV., supra note 4, at 27 (“Data-driven innovation] is a disruptive new source of growth that could transform all sectors in the economy.”).

132. See sources cited supra note 8.

133. As Keith Maskus noted:

That data have economic value at the macroeconomic level is mirrored by the fact that they are intensely valuable for firms that generate or collect them online. This value stems largely from the ability of firms to monetise data through selling their characteristics to advertisers. It is estimated that Google’s advertising revenue per user was US$ 45 in early 2014 and US$ 9.45 for Facebook in the same year. Each of these companies earned more than 90 percent of its revenue from online advertising. Access to data also permits firms to tailor new products and services to
the considerations for the data producer's right may be significantly different from those earlier ones for *sui generis* database protection. Moreover, it is one thing to lose a competitive edge in the database industry, but quite another thing to lose a competitive edge in a multitude of data-driven industries.

Notwithstanding the enormous value in data and their considerable economic significance, it remains debatable whether new rights should be introduced to protect machine-generated data in the first place. After all, many of the targeted data will already be generated regardless of the existence of these new rights. As Josef Drexl explained, using the example of data generated by the networked sensors in a smart car:

The data produced . . . will be transferred to the manufacturer of that car. The car manufacturer will be sufficiently motivated to generate data that will guarantee smooth operation and maintenance of the car. Generation of that data is very much part of the firm's business model . . .

Nor are additional incentives needed as regards the business model of Internet platform operators (e.g., search engines, social media etc.), for which the collection of personal data is the very core of the success of the underlying business model.

In addition, the benefits provided by the introduction of a data producer's right may not outweigh the costs incurred by this particular right. To help us think through the complex cost-benefit analyses and meet the preferences of individual users, raising the value of data even more. The current trend of massive private investments made in constructing databases is additional evidence of the large economic stakes in this sector. As for the EU, it is estimated that applications built on personalised data could provide benefits to its firms and citizens of perhaps €1 trillion annually by 2020.

Maskus, *supra* note 129, at 23 (footnotes omitted).

134. *See* Drexl, *supra* note 17, at 291 ("In principle, in the data economy, no incentives are needed for generating and commercialising data."); Kerber, *supra* note 5, at 997 ("Both from a theoretical and empirical perspective there is no evidence that there are generally too few incentives for producing and analyzing data in the digital economy . . . Without an incentive problem the main economic argument for establishing an exclusive property right vanishes, if the use of the good is simultaneously non-rivalrous (as in the case of data."); *id.* at 990 ("[O]n the basis of our current preliminary knowledge—a new [intellectual property right] on data is not necessary (especially due to the lack of an incentive problem for producing and analyzing data.")."

135. Drexl, *supra* note 17, at 273; *see also* P. Bernt Hugenholtz, *Data Property in the System of Intellectual Property Law: Welcome Guest or Misfit?*, in *TRADING DATA IN THE DIGITAL ECONOMY: LEGAL CONCEPTS AND TOOLS* 75, 80 (Sebastian Lohse et al. eds., 2017) ("Much machine data production occurs (nearly) automatically, often as a by-product of industrial production or services, and it is hard to see why a legal incentive in the form of a data property right would enhance it.").
the multiple trade-offs involved, this Part explores whether the proposed data producer’s right would meet the needs of our present technological environment. This Part further explores the proposed right in light of the needs of the U.S. business and scientific communities and society at large.

A. Technological Needs

Although the increased use of big data analytics and the proliferation of Internet-of-Things devices has made machine-generated data highly important, many counterarguments exist to caution against the introduction of a new data producer’s right. One particularly strong argument concerns how good big data analyses require the existence of large, comprehensive datasets. If Stewart Brand’s famous quote about how information wants to be both free and expensive can be modified in the big data context, that quote would read, “Data want to be expensive, but they also want to be complete.”

Moreover, because of the changing nature of our technological environment, many relevant data now reside in separate datasets and often in multiple data storage systems. In the past decade, computer scientists and engineers have worked tirelessly to develop ways to

---

136. See Maskus, supra note 129, at 21 (noting the “multiple trade-offs” involved in developing policy frameworks to promote emerging industries and new innovation).

137. See Mayer-Schönberger & Cukier, supra note 1, at 30 (noting that “big data relies on all the information, or at least as much as possible”).

138. The original quote is as follows:

Information wants to be free because it has become so cheap to distribute, copy, and recombine—too cheap to meter. It wants to be expensive because it can be immeasurably valuable to the recipient. That tension will not go away. It leads to endless wrenching debate about price, copyright, “intellectual property,” and the moral rightness of casual distribution, because each round of new devices makes the tension worse, not better.


139. See Manyika et al., supra note 4, at 12 (“To enable transformative opportunities, companies will increasingly need to integrate information from multiple data sources.”); Mayer-Schönberger & Cukier, supra note 1, at 46 (“Large datasets do not exist in any one place; they tend to be split up across multiple hard drives and computers.”); Riley, supra note 7, at 254 (“One of the biggest challenges for Big Data [in the health care space] is linking data from multiple sources so that data describing an individual located in one source are linked with data about the same individual in other sources.”).
analyze data without moving them from one storage system to another.\textsuperscript{140} Thus, if we are to maximize our ability to undertake big data analyses, such analyses may require greater sharing of data—which, in turn, calls for greater data portability and interoperability.\textsuperscript{141}

Indeed, allocating property rights to data producers—both individual and corporate—could create major market barriers, leading to what Rebecca Eisenberg and Michael Heller have described as the “tragedy of the anticommons.”\textsuperscript{142} Instead of facilitating socially productive uses of machine-generated data, the proposed data producer’s right would fragment the data market, undermining the benefits of new, innovative data analytical techniques.\textsuperscript{143}

\textbf{B. Business Needs}

At the time of writing, U.S. companies dominate the global data-driven economy.\textsuperscript{144} Such dominance was indeed why many European policy makers were so eager to introduce a new data producer’s right.\textsuperscript{145}

\begin{itemize}
\item \textsuperscript{140} See John D. Kelleher & Brendan Tierney, \textit{Data Science} 78-80 (2018) (discussing Hadoop and other efforts to move the algorithms to the data, as opposed to moving the data themselves); President’s Council of Advisors on Sci. & Tech., supra note 23, at 30 (“Specialized software technology allows the data in multiple data centers (and spread across tens of thousands of processors and hard-disk drives) to cooperate in performing the tasks of data analytics, thereby providing both scaling and better performance.”).
\item \textsuperscript{141} See discussion infra Part IV.A.5.
\item \textsuperscript{143} See Drexdl, supra note 17, at 260 & n.16 (considering “multiple ownership of the same data with considerable negative effects on access to that data” as “a situation of a ‘tragedy of the anti-commons’ in which too many property rights in the same asset lead to inefficient underuse of that asset”); Kerber, supra note 5, at 990 (“The introduction of a new intellectual property right in data can be ... dangerous for innovation and competition in the digital economy, because it might lead to considerable legal uncertainty, the monopolisation of information, and impediments for the free flow of data that is so crucial for the digital economy.”).
\item \textsuperscript{144} See Reto Hilty, \textit{Big Data: Ownership and Use in the Digital Age}, in \textit{Intellectual Property and Digital Trade}, supra note 15, at 85, 89 (“The majority of the dominating drivers of the digital economy are indeed US companies (the Googles, Facebooks, WhatsApps, Amazons, and many others). And it is certainly true that those companies would not be as successful as they are if they did not have access to ‘big data’, or, more precisely, to our data.” (emphasis omitted)).
\item \textsuperscript{145} See Hugenholtz, supra note 6, at 49 (“Apparently inspiring this call for protecting industrial data is the fear—common to other recent policy initiatives—that valuable European assets are being misappropriated by large American companies.”).
\end{itemize}
These policy makers hoped that the new right would help slow down the competition from American technology giants, such as Google, Apple, Facebook, and Amazon, which European commentators and critics have lumped together with the acronym “GAFA.”

Looking back, the EU policy makers’ eagerness to out-compete the United States was a primary reason for adopting the EU Database Directive. Recital 11 of the Directive states that “there is at present a very great imbalance in the level of investment in the database sector both as between the Member States and between the Community and the world’s largest database-producing third countries.” Recital 12 further declares that “investment in modern information storage and processing systems will not take place within the Community unless a stable and uniform legal protection regime is introduced for the protection of the rights of makers of databases.”

Thus far, U.S. technology companies have greatly benefited from the free flow of data. While the proposal for a new data producer’s right could provide some benefits to these companies, due to their ability to produce a prodigious quantity of data, that same proposal would also stifle corporate technological development. It is no wonder that these for-profit companies did not lobby for stronger protection of machine-generated data. Their refusal to do so may have indicated their belief that the free flow of data would, on balance, provide more overall benefits than greater data protection and a further reduction of data flows.


147. See Hugenholtz, supra note 6, at 49 (“The sui generis database producer’s right introduced in Europe in 1996 was similarly inspired by European fears of dominance by the US database industry.”). As Reichman and Samuelson recounted:

The Commission found that European database producers had to overcome several comparative disadvantages in order to expand their share of the world market and to catch up with the U.S. industry, which dominated the market and was growing at a faster rate than its European counterpart. To overcome these disadvantages, the Commission stressed the need for a single, integrated market, undistorted by differing regulatory approaches, and for higher levels of intellectual property protection, tailored to the needs of potential investors in database production, which might stimulate additional investment in this sector.


149. Id. recital 12.

150. See Drexl, supra note 17, at 260 (“[M]any firms are producers of data and have to rely on access to data of other players at the same time.”).

151. As Drexl suggested:
Given the preference of leading U.S. technology companies, it is understandable why U.S. negotiators have been actively negotiating for language to foster the free flow of data at the global level. Cases in point are the electronic commerce chapters found in bilateral, regional, and plurilateral trade agreements,\textsuperscript{152} which the United States has negotiated since the early 2000s.\textsuperscript{153} Part IV.B.3 will discuss these agreements in greater detail.

Data analysts will not gain a competitive edge by “owning data” at the expense of their competitors. Rather, they will prevail in competition if they manage to have better access to the various sources of big data, for which they will not rely on ownership but contractual business relationships with the holders of such datasets, on the one hand, and the effectiveness and accuracy of their big data analyses, on the other hand. As regards the latter, it is more important that big data analysts control the technology for big data analysis. For this, they will rely on copyright protection in the software infrastructure and possibly technical know-how rather than data ownership. The same holds true for firms that deliver—typically software-based—tools for big data analysis of other firms.

\textit{Id.} at 274 (footnote omitted); see also \textit{Ajay Agrawal et al., Prediction Machines: The Simple Economics of Artificial Intelligence} 189 (2018) (noting that technology companies understand that learning in the artificial intelligence space “often requires customers who are willing to provide data”).

152. As the Office of the U.S. Trade Representative noted in its summary of the TPP electronic commerce chapter:

\textit{[This] chapter includes commitments ensuring that companies and consumers can access and move data freely (subject to safeguards, such as for privacy), which will help ensure free flow of the global information and data that drive the Internet and the digital economy. These commitments, along with others on market access and national treatment, combine to help prevent unreasonable restriction, such as the arbitrary blocking of websites.}

\textit{...[D]ata flows are the building blocks for all digital trade, and barriers to them are among the most serious impediments to the future of digital trade. Impediments to such flows affect not only technology companies, but almost every sector of the economy from manufacturing to farming and small businesses—all of which now depend on digital technology to provide the innovation and efficiency that drive economic growth.}


Moreover, if we can draw lessons from the EU Database Directive, a key lesson relating to this subpart is that well-intentioned efforts to improve global competitiveness through data protection does not always succeed. Indeed, commentators readily admit that the adoption of the EU Database Directive was a mistake. 154 Considering that the current global norms already favor U.S. technology companies, one has to question the wisdom of gambling on a new policy that may or may not benefit these companies.

C. Scientific Needs

Like the EU Database Directive and U.S. sui generis database legislation, the proposal for a new data producer’s right could undermine the ongoing effort by the research and library communities to share data with others through open science and open data initiatives. 155 The proposed right would create both actual and transaction costs." As Jerome Reichman, Paul Uhlir, and Tom Dedeurwaerdere noted in their recent book:

[1] Inter-university exchanges of data are increasingly subject to high transaction costs, delays, and a growing risk of anti-commons effects, that is, too many intellectual property rights and commercial interests making it difficult to build comprehensive or complex databases. This problem is particularly acute in cases of transnational scientific collaboration. As relations between universities and industry become more intense, and as public universities receive a smaller share of their budgets from state legislatures, the universities tend to view each other


154. See Davison, supra note 40, at 829 ("It is difficult to draw any conclusion other than that the adoption of the [EU Database] Directive was a mistake.").

155. See Commission Staff Working Document, supra note 10, at 36 (noting “the [European] Commission’s policy on open science and open access”); Mayer-Schönberger & Cukier, supra note 1, at 116-18 (discussing the value of open data and highlighting the open data initiatives around the globe); Reichman et al., supra note 8, at 367 ("[A] number of public or publicly funded organizations have decided to make their data holdings openly available.").

156. As Jerome Reichman and Paul Uhlir observed in the late 1990s:

[U]nder the current [sui generis database protection] proposals, scientists and engineers would face rising transaction costs when attempting to create complex databases from multiple public and private sources. Also predicted are higher costs due to the burdens of administering national data centers and of carrying out related, large-scale management activities that currently benefit from the policy of open and unrestricted access to scientific and technical data.

Reichman & Uhlir, supra note 61, at 816.
as competitors, rather than as partners in a common mission. Their industrial partners are correspondingly more likely to impose their own proprietary terms of exchange on the universities. 157

To a large extent, some of these concerns resemble those earlier concerns that commentators expressed when Congress was actively considering *sui generis* database protection bills two decades ago. 158 Compared with that time, the present stakes are much higher, especially when one considers the myriad benefits provided by the free exchange of data for scientific pursuits and the growing success of open science and open data initiatives. If the development of *sui generis* protection for databases was already ill-advised at that time, granting *sui generis* protection to machine-generated data seems clearly wrongheaded.

D. Personal Needs

The increasing use of data has created considerable concern about how data—whether personal or anonymized—are being used. The arrival of new Internet-of-Things devices has also led to the development of what Frank Pasquale called the "black box society." 159 As he observed, "Black boxes embody a paradox of the so-called information age: Data is becoming staggering in its breadth and depth, yet often the information most important to us is out of our reach, available only to insiders." 160 Using popular services such as Google, Facebook, and Twitter as illustrations, he elaborated:

Without knowing what Google actually *does* when it ranks sites, we cannot assess when it is acting in good faith to help users, and when it is biasing results to favor its own commercial interests. The same goes for status updates on Facebook, trending topics on Twitter, and even network management practices at telephone and cable companies. All these are protected by laws of secrecy and technologies of obfuscation. 161

158. *See* sources cited *supra* note 80.
160. *Pasquale, supra* note 159, at 191; *see also* Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42-43 (2013) ("Big data promises to use . . . data to make the world more transparent, but its collection is invisible, and its tools and techniques are opaque, shrouded by layers of physical, legal, and technical privacy by design.").
161. *Pasquale, supra* note 159, at 9; *see also* Kent R. Anderson, *Can We Anticipate Some Unintended Consequences of Big Data?, in Big Data Is Not a Monolith, supra* note
Although most of the ongoing discussion in what Pasquale described as the “era of runaway data”\textsuperscript{162} has thus far focused on personal data, it is unclear when seemingly anonymized machine-generated data would become personal or personally identifying. As one commentator observed, “Even when aggregating non-personal data (e.g. machine data from a board computer), powerful correlation algorithms working on big data may be able to correlate such data with an individual person, thus transforming non-personal data into personal data.”\textsuperscript{163} Likewise, Viktor Mayer-Schönberger and Kenneth Cukier lamented, “Given enough data, perfect anonymization is impossible no matter how hard one tries.”\textsuperscript{164} Their book showed disturbingly how the New York Times staff and researchers at the University of Texas at Austin were able to reconstruct seemingly anonymized data from companies such as AOL and Netflix.\textsuperscript{165}

To some extent, the increased demands for transparency in relation to these “black box” devices resonate with the growing demands for transparency in international intellectual property negotiations. The latter demands were shown vividly by the protests\textsuperscript{166} against controversial U.S. legislation,\textsuperscript{167} the Anti-Counterfeiting Trade

\begin{footnotesize}
\textsuperscript{7} at 117, 124 (“The data in Google flu trends is presented as is, and cannot be analyzed by outside parties.”).

\textsuperscript{162} PASQUALE, supra note 159, at 19.

\textsuperscript{163} Peter Bitiner, Intellectual Property Management Challenges Arising from Pervasive Digitalisation: The Effect of the Digital Transformation on Daily Life, in INTELLECTUAL PROPERTY AND DIGITAL TRADE, supra note 15, at 67, 71; see also ARI EZRA WALDMAN, PRIVACY AS TRUST: INFORMATION PRIVACY FOR AN INFORMATION AGE 65 (2018) (“Any individual piece of information may not be particularly intimate or personal, but its sum total can paint a detailed picture.”).

\textsuperscript{164} MAYER-SCHÖNBERGER & CUKIER, supra note 1, at 155.

\textsuperscript{165} See id. at 154-55 (recounting these incidents).


\end{footnotesize}
DATA PRODUCER'S RIGHT

Agreement, and the Trans-Pacific Partnership (TPP). Both sets of demands reflect changing consumer expectations. In the mid-1990s, when Congress was actively considering sui generis database protection bills to match the EU Database Directive, Boyle issued a pioneering call for the creation of “a politics of intellectual property.” Two decades later, however, these politics have slowly emerged, and the demands for greater transparency now inform the debate on issues ranging from international intellectual property negotiations to the construction or design of Internet-of-Things devices.

E. Summary

This Part has shown that while data have become highly valuable and greater protection could help lock in their enormous value, the proposal for a new data producer’s right is misguided. That proposal does not meet the present technological, business, scientific, and personal needs, especially those in the United States. Given this assessment, it is doubtful that the proposed data producer’s right would be, on balance, socially expedient. Even if the proposed right would


172. As Neil Richards and Jonathan King observed: [W]e must recognize that big data requires transparency. Transparency has long been a cornerstone of civil society as it enables informed decision making by governments, institutions, and individuals alike. The many secondary uses of big data analytics, and the resulting incentives of companies and governments to share data, place heightened importance on transparency in our age of big data. Transparency can help prevent abuses of institutional power while also encouraging individuals to feel safe in sharing more relevant data to make better big data predictions for our society.

Neil M. Richards & Jonathan H. King, Big Data Ethics, 49 WAKE FOREST L. REV. 393, 396 (2014); see also Cate, supra note 15, at 18 (“Whenever big data are used in ways that affect individuals, there must be effective transparency and redress.”).
provide important incentives to facilitate the production of machine-generated data, such benefits might not compensate for the many costs this new right would incur.

IV. FUTURE DEVELOPMENTS

Drawing on past and present developments, the previous two Parts have shown why the proposal for a new data producer's right is ill-advised, especially when introduced in the United States. This Part turns to the complications the proposed right would create in relation to the future development of a sound and holistic data governance regime. This Part breaks these complications down into two subgroups, based on whether they are within or outside the area of machine-generated data.

Focusing on endogenous complications, subpart A points out that the proposed data producer's right would create more questions than answers. If policy makers strongly believe that such a right is needed to provide incentives for the production of machine-generated data, they will have to provide satisfactory answers to these difficult questions.

Subpart B then turns to exogenous complications. This subpart shows that the proposal for a new data producer's right would create considerable complications not only in the intellectual property system but also in other international regulatory systems, such as those governing privacy, trade, and investment. The awareness of these complications is highly important because the issues are increasingly connected, especially when policy makers explore ways to develop a sound and holistic data governance regime.

A. Endogenous Complications

Thus far, commentators have provided different justifications for the introduction of a new data producer’s right. As Herbert Zech, a professor at the University of Basel in Switzerland and the right's leading proponent,173 explained:

[A]s in classic intellectual property rights, one could bring forward the argument that incentives are created to generate and to reveal data (and hence, indirectly, to promote innovations that are made possible through

the use of data) and that markets for information goods (that otherwise would not be tradable or would only be tradable with higher transaction costs) are created. Another important aspect seems to be that such legal regulation would clearly determine who benefits from the use of data. This would prevent that machines are designed in a way that they are difficult to be read out or that other mechanisms would be created which grant de facto exclusivity. Such a regulation not only saves costs, but would promote a culture of transparency, as “open data” does. The data producer right would have the same function for “open data” as the copyright has for “open source” and “open content”.174

Given these justifications, the European Commission’s eagerness to offer protection to machine-generated data is indeed not difficult to explain, even if one does not take into account the heavy lobbying by the automotive industry.175 After all, uncertainty over the rights involved in the data environment could discourage investment in data-related innovations.176 Such uncertainty could also create disputes, especially in view of the fast-increasing value in, and multifaceted uses of, machine-generated data.177

Notwithstanding the need to address these understandable concerns, the proposal for a new data producer’s right has raised many difficult policy questions that have to be addressed before the creation of this new right. Even if one could provide a fairly strong justification for this right—such as when its creation is needed to encourage data producers to disseminate or share machine-generated data for socially

174. Id.

175. See Hugenholtz, supra note 6, at 48 (noting the “demands of the automotive industry”).

176. See Commission Communication, supra note 2, at 13 (noting that the creation of a new data producer’s right “would aim at clarifying the legal situation”); Zech, supra note 173, at 470 (“[A]n important aspect [of the data producer’s right] seems to be that such legal regulation would clearly determine who benefits from the use of data.”). But see Drexl, supra note 17, at 275 (“As regards data ownership that is recognised independently of factual control over data in an environment where individual data may constantly be integrated and arranged in different datasets, data ownership is more likely to reduce transparency and increase the risk of unintentional infringement of rights.”).

177. See Drexl, supra note 17, at 275 (“[N]ew property rights will always give rise to additional conflicts and litigation.”); Timo Minssen & Justin Pierce, Big Data and Intellectual Property Rights in the Health and Life Sciences, in BIG DATA, HEALTH LAW, AND BIOETHICS, supra note 7, at 311, 311 (noting the “fierce controversies between data ‘owners,’ data researchers, and entities that provide enabling technologies, large research infrastructures, and standardization platforms”).
beneficial purposes\textsuperscript{178}—answers to these questions will still be needed. This subpart discusses each set of policy questions in turn.

1. Modality of Protection

The first set of policy questions concerns the modality of protection. As far as intellectual property laws are concerned, policy makers and commentators tend to create new rights utilizing the property model. Despite its many strengths, this model is not always the best for developing intellectual property protection.\textsuperscript{179} Nor is it the only model for developing such protection. As Part I.I.A noted, the interests of data producers could be easily protected using trade secrecy, unfair competition law, cost-sharing arrangements, private contracts, or technological protection measures.\textsuperscript{180} One should not forget that the original proposal for the EU Database Directive was based on unfair competition law principles.\textsuperscript{181}

\textsuperscript{178} This quid-pro-quo arrangement is generally used to justify patent protection. Without such protection, many inventions may remain locked up in secrecy. See Edith Tilton Penrose, The Economics of the International Patent System 31-34 (1951) (discussing the disclosure of secrets as a rationale of the patent system); see also A. Samuel Oddi, The International Patent System and Third World Development: Reality or Myth?, 1987 Duke L.J. 831, 851 ("the argument most commonly made is that a patent incentive is needed to induce the transfer of technology because patent owners would otherwise be unwilling to transfer their valuable technology in the form of trade secrets due to the relative weakness of trade secret law in developing countries.").

\textsuperscript{179} See Yu, supra note 29, at 792 ("As far as policy options are concerned, there is a misguided tendency for policymakers in both developed and less developed countries to assume that the property rights model is the only model, or the best one, that is compliant with the TRIPs Agreement or other commitments under the international intellectual property regime."); Peter K. Yu, Ten Common Questions About Intellectual Property and Human Rights, 23 Ga. St. U. L. Rev. 709, 735 (2007) ("[A]lthough a property-based intellectual property system would offer the needed protection to material interests in intellectual creations, such a regime is not the only acceptable, or even the best, modality of protection that can be used to realize the right to the protection of material interests in intellectual creations.").

\textsuperscript{180} See Drexel, supra note 17, at 292 ("As regards access negotiations between private parties, the Commission could support schemes of private ordering that enable private initiatives to pool data of multiple data holders."); Hugenholtz, supra note 6, at 52 ("[N]on-property regimes such as contracts and trade secret protection might occasionally do the job, as would technical protection measures that create de facto ownership positions."); see also supra notes 82-85 and accompanying text.

\textsuperscript{181} As Reichman and Samuelson recounted: The Commission’s initial approach was premised on the absence of a harmonized system of unfair competition legislation to safeguard “the investment of considerable human, technical and financial resources” in the making of databases that “can be copied . . . at a fraction of the cost needed to design them independently.” The logical solution was, therefore, to codify a new type of unfair competition law. Such a law, loosely modelled on existing laws that protected trade secrets or confidential information, would repress conduct amounting to the
To be sure, a data producer’s right established through trade secrecy will be very different from one utilizing the property model. While the latter prevents any party from using the generated data absent any permissible limitation or exception, the former requires data producers to undertake reasonable precautions to ensure that the data remain protected through secrecy. Likewise, although contracts offer protection to data producers, such protection is limited to those enjoying the privity of contract.

“misappropriation” of an electronic database producer’s investment without imposing either legal barriers to entry or the social costs of actual or legal secrecy. To this end, the Commission proposed simply to forbid the “unfair extraction” of data from an electronic database for commercial purposes without the second comer’s having expended independent effort to collect and verify similar information. The first proposed draft E.C. Directive accordingly provided a ten-year period of lead time in which the database maker could recoup his or her investment in a noncopyrightable electronic database while preventing copiers from engaging in for-profit extraction or reutilization of the factual contents, in whole or in substantial part.


182. As Drexl explained:

[Although the Trade Secrets Directive was not drafted to meet the needs of the data economy, trade secrets protection can provide a sound approach to protecting firms in the data economy to some extent. Rather than recognising exclusive control over any use of protected information, as would be typical for intellectual property regimes, EU trade secrets law implements a tort law approach that bans specific conduct related to the acquisition, dissemination and use of trade secrets that can be considered as unfair. It is thereby better suited to balance the interest in protection and in free flow of information than the property approach.]

Drexl, supra note 17, at 291. But see Commission Staff Working Document, supra note 10, at 20 (“It is doubtful that individual data generated by interconnected machines and devices could be regarded as ‘trade secret’ in the sense of this Directive, mostly because of its lack of commercial value as individual data . . . .”).

183. See Council Directive 2016/943, art. 2(1)(c), 2016 O.J. (L 157) 1 (requiring a trade secret to have been “subject to reasonable steps under the circumstances, by the person lawfully in control of the information, to keep it secret”); UNIF. TRADE SECRETS ACT § 1(4) (amended 1985), 14 U.L.A. 536 (2005) (requiring a trade secret to have been “the subject of efforts that are reasonable under the circumstances to maintain its secrecy”). See generally Robert G. Bone, Trade Secrecy, Innovation and the Requirement of Reasonable Secrecy Precautions, in THE LAW AND THEORY OF TRADE SECRECY: A HANDBOOK OF CONTEMPORARY RESEARCH 46 (Rochelle C. Dreyfuss & Katherine J. Strandburg eds., 2011) (discussing the need for precautionary measures).

184. See ProCD, Inc. v. Zeidenberg, 86 F.3d 1447, 1454 (7th Cir. 1996) (“Contracts . . . generally affect only their parties; strangers may do as they please . . . .”); NAT’L RESEARCH COUNCIL, supra note 69, at 61 (“The fact that contract terms are only effective between the contracting parties and not binding on third parties who may get access to the database has been cited as a weakness, since many databases must be publicly distributed in order to be commercially viable.” (footnote omitted)); Reichman & Samuelson, supra note 43, at 137
The different protections offered by these rights were indeed the focus of the European Commission. As the Commission stated in its staff working document, the proposed right "could be envisaged as a right in rem" or "as a set of purely defensive rights."\(^{185}\) As the Commission elaborated:

[The data producer's] right could be envisaged as a right in rem and assign the exclusive right to utilise certain data, including the right to licence its usage. This would include a set of rights enforceable against any party independent of contractual relations thus preventing further use of data by third parties who have no right to use the data, including the right to claim damages for unauthorised access to and use of data.

....

Alternatively, instead of creating the data producer right as a right in rem, it could be conceived of as a set of purely defensive rights. This option would follow the choice made in the design of the protection given to know-how by the Trade Secrets Protection Directive. Its objective would be to enhance the sharing of data by giving at least the defensive elements of an in rem right, i.e. the capacity for the de facto data holder to sue third parties in case of illicit misappropriation of data. This approach thus equates to a protection of a de facto "possession" rather than to the concept of "ownership."\(^{186}\)

The Commission's explanation is particularly instructive on issues relating to the modality of protection. While the "right in rem" discussion covers the strengths of the property model, the "defensive rights" approach steers the protection toward trade secrecy or unfair competition laws.\(^{187}\)

\(^{185}\) Commission Staff Working Document, supra note 10, at 33 (emphasis omitted).
\(^{186}\) Id. at 33-34 (footnotes and emphasis omitted).
\(^{187}\) As Reto Hilty noted: [The "defensive rights"] approach resembles "possession" rather than "ownership;" it is comparable to the possession of (as such not protected) know-how, and the concept of legal protection may be similar to the one applied in Directive 2016/943 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.

Hilty, supra note 144, at 89.
2. Allocation of Rights

The second set of policy questions relates to allocation of rights. If the property model is to be used, such allocation will focus on ownership interests. Thus far, three groups of players have fairly strong claims to machine-generated data, even though one could arguably include other groups in the mix. In his proposal for a data producer’s right, Zech described a data producer as “the economically responsible operator of equipment that generates the data.” This position contrasts interestingly with extant European copyright laws that consider those making necessary arrangements for the creation of computer-generated works as authors.

For illustrative purposes, this subpart considers the data produced by the networked sensors in a smart car—an example that has been used ad nauseam in discussions of machine-generated data. Most

188. See Farkas, supra note 10, at 6-7 (suggesting as potential owners of the data produced by the networked sensors in a smart car “[t]he owner of the car, the actual user of the car, the car manufacturer, the manufacturer of the sensor or communicating devices built in the networked car, companies like navigation service providers [and] the road construction authorities”); Hilty, supra note 144, at 91 (“[I]n the case of the traffic app, who should be the ‘owner?’ Should it be the car producer, the supplier of the sensor or control unit, the app producer, the service provider, the car driver—or even perhaps another party?”); see also Drexl, supra note 17, at 277 (“[I]n a complex world of networks where a considerable number of different players collaborate in generating value, not least by contributing their data, the allocation of data ownership is particularly difficult.”); Kerber, supra note 5, at 991 (“An especially difficult problem seems to be the question to whom such an [intellectual property right] should be granted: Should it be the data producer, who codifies the data . . . , or the firm which is economically responsible for the production of data, or the firm which can benefit most from the data?”).

189. Zech, supra note 173, at 469. In its staff working document, the European Commission defined this term more broadly:

One of the criteria for allocating the right could be to take into account the investments done and the resources put into the creation of the data. Such investments are made most often by two sides: The manufacturer of sensor-equipped machines, tools or devices (generating the data) who has invested in to the development and market commercialisation of the machine, tool or device and the economic operators using such machines, tools or devices paying a purchase price or lease and have to amortise the machine, tool or device.

Commission Staff Working Document, supra note 10, at 35.

190. See, e.g., Copyright, Designs and Patents Act 1988, c. 48, § 9(3) (Eng.) (“In the case of a literary, dramatic, musical or artistic work which is computer-generated, the author shall be taken to be the person by whom the arrangements necessary for the creation of the work are undertaken.”); Copyright and Related Rights Act 2000 (Act No. 28/2000) § 21(f) (Ir.), http://www.irishtatutebook.ie/eli/2000/act/28/enacted/en/html (stipulating that the author of a computer-generated work shall be “the person by whom the arrangements necessary for the creation of the work are undertaken”).

191. See, e.g., Commission Staff Working Document, supra note 10, at 25 (“Access to and re-use of non-personal or anonymised data have been subject to intensive discussions with
certainly, the car owner has a strong claim to these data, as the owner will need them to repair the car.192 For privacy protection, the owner may also need the ability to opt out from automatic data generation, especially if he or she does not trust the anonymization process.193 The owner's demand for opting out is easy to understand. While the produced or collected car data may provide valuable information about traffic patterns, road conditions, and engine performance,194 they could also reveal highly personal information about the owner's driving habits.195

Like the car owner, the car manufacturer also has a strong claim to these data, as the manufacturer will need them to improve car design, anticipate problems generated by design defects, and develop solutions to address those problems.196 In recent years, some car manufacturers


192. See Mayer-Schönberger & Cukier, supra note 1, at 133 (“Cars today are stuffed with chips, sensors, and software that upload performance data to the carmakers’ computers when the vehicle is serviced.”); Drexl, supra note 17, at 262 (“[S]mart cars nowadays collect data . . . for providing better and timely—even predictive—maintenance services.”); Farkas, supra note 10, at 6 (“A networked car might communicate details about traffic, favourite routes and road conditions to the car owner . . .”); see also Commission Staff Working Document, supra note 10, at 25 (discussing the handling of data from connected vehicles).

193. Cf. Commission Communication, supra note 2, at 13 (“Where personal data are concerned, the individual will retain his right to withdraw his consent at any time after authorising the use.”). Nevertheless, such opting out may not always be easy, given that some of the produced or collected data are essential to the car's safe and effective operation.

194. See Hugenholtz, supra note 6, at 48-49 (noting the observation of European Commissioner for the Digital Economy and Society Günther Oettinger that “modern sensor-equipped cars automatically generate and collect large amounts of data—on traffic and road conditions, engine performance, etc.”).

195. See Drexl, supra note 17, at 262 (noting that “cars may . . . register the driving habits of the driver”); Farkas, supra note 10, at 6 (noting that the data communicated by networked cars “can be used to avoid . . . driving behaviour relevant for insurers”).

196. As the Commission explained in its staff working document:

In case the right is allocated to the economic operator using the machine, tool or device, exceptions may need to be made for the manufacturer of such machine, tool or device. The manufacturer may not only have a legitimate interest to use such data for the purposes of further improving product design, but also may have a legal obligation to monitor the behaviour of his products on the market. In some circumstances, there may be good reasons not to allocate the ownership to the economic operator using the machine, tool or device—or not to allow the full range of actions usually available to an “owner”—notably on safety or security linked to the relevant data.

Commission Staff Working Document, supra note 10, at 35-36 (footnote omitted). Likewise, Michael Bailey observed:
have suggested, somewhat counterintuitively, that car owners do not actually have ownership interests despite paying handsomely for their vehicles. If these owners did not even own the cars they bought, it would be highly unlikely for them to own the data generated by the sensors in those cars.

As if the fight over data control between these two parties were not complicated enough, the software developer has an equally strong claim to the data produced by these sensors. Just as the car owner and the car manufacturer need data to repair or improve the car, the software designer also needs data to improve the software.

Big data is changing how automakers and firms do recalls. Firms can now store many more data about how products are produced, sold, and delivered, and link these to all reported problems and complaints, allowing companies to identify the source of problems and connect related problems. General Motors was able to limit the recall of a certain model of car with a faulty valve to only four cars because of its advanced manufacturing tracking.

This suggestion occurred in the U.S. Copyright Office’s 2015 rulemaking proceeding concerning the anti-circumvention provision of the DMCA. See Peter K. Yu, The Anatomy of the Human Rights Framework for Intellectual Property, 69 SMU L. REV. 37, 89 (2016). As I recounted in an earlier article:

In its submission arguing against the introduction of exceptions in Class 21 for “vehicle software—diagnosis, repair, or modification,” John Deere claimed that those who purchased its tractors did not have ownership interests in those vehicles; instead, they merely “receive[d] an implied license for the life of the vehicle to operate the vehicle, subject to any warranty limitations, disclaimers or other contractual limitations in the sales contract or documentation.” Automobile manufacturers also made similar arguments supporting the post-sale control of cars they produced. As General Motors stated in its submission, those who own its cars do not own the computer software in the vehicles even though such software, it admits, is essential to the vehicle’s safe operation.


See Drexl, supra note 17, at 267 (“The property in the car as a physical object does not automatically extend to the commercial exploitation of the data that are produced by the sensors of that car.”).

See Paul R. Daugherty & H. James Wilson, Human + Machine: Reimagining Work in the Age of AI 114 (2018) (noting the use of data and artificial intelligence systems for “diagnosing and fixing problems with [information technology] systems”); Mayer-Schonberger & Cukier, supra note 1, at 105 (noting the potential use of voice-translation records to improve speech-recognition technology); id. at 107 (noting that mobile phone “[o]perators have long used [the data about where and when the phones connect to base
Moreover, the programmers hired by this developer have designed the software to enable the sensors and the car computer to produce and collect data. In cases involving copyright law and artificial intelligence, commentators have widely agreed that computer programmers should have ownership interests in computer-generated works. As Annemarie Bridy observed, "Intuition and the principle of transitivity both suggest that the programmer of generative software is the logical owner of the copyright in the works generated by his or her software. He or she is, after all, the author of the author of the works."  

Nevertheless, if considerable progress has been made for machines to conduct "deep learning" or "artificial general intelligence" to fine-tune the performance of their networks, deciding where to add or upgrade infrastructure; REICHMAN ET AL., supra note 8, at 323 ("One of the most promising user-added resources resulting from data accumulation and integration is the establishment of incremental machine learning and automatic interpretation capabilities based on the application of semantic web and integrative techniques to large amounts of data.").  

200. See, e.g., NAT’L COMM’N ON NEW TECH. USES OF COPYRIGHTED WORKS, FINAL REPORT OF THE NATIONAL COMMISSION ON NEW TECHNOLOGICAL USES OF COPYRIGHTED WORKS 45 (1979) ("[W]e confront the question of who is the author of a work produced through the use of a computer. The obvious answer is that the author is [the] one who employs the computer."); Annemarie Bridy, Coding Creativity: Copyright and the Artificially Intelligent Author, 2012 STAN. TECH. L. REV. 5, 21 (finding "the programmer of generative software . . . the logical owner of the copyright in [computer-generated] works"); Arthur R. Miller, Copyright Protection for Computer Programs, Databases, and Computer-Generated Works: Is Anything New Since CONTU?, 106 HARV. L. REV. 977, 1049 (1993) ("[T]he human element in the creation of [computer-generated] works is sufficient to sustain their copyrightability and resolve any question of authorship."); Pamela Samuelson, Allocating Ownership Rights in Computer-Generated Works, 47 U. PITTSB. L. REV. 1185, 1192 (1986) ("In general, the user of a computer generator program should be considered the author of a computer-generated work, and should be free to exploit this product commercially."); see also Kal Raustiala & Christopher Jon Sprigman, The Second Digital Disruption: Streaming & the Dawn of Data-Driven Creativity, 94 NYU L. REV. (forthcoming 2019) (manuscript at 65), https://ssrn.com/abstract=3226566 ("[E]ven in a world where data-driven authorship is the norm, we would still recognize that authors are laboring—at least where authors are assisted by data and algorithms, rather than being entirely displaced by them.").  

201. Bridy, supra note 200, at 21.  

202. As a government report on artificial intelligence observed:

In recent years, some of the most impressive advancements in machine learning have been in the subfield of deep learning, also known as deep network learning. Deep learning uses structures loosely inspired by the human brain, consisting of a set of units (or "neurons"). Each unit combines a set of input values to produce an output value, which in turn is passed on to other neurons downstream. For example, in an image recognition application, a first layer of units might combine the raw data of the image to recognize simple patterns in the image; a second layer of units might combine the results of the first layer to recognize patterns-of-patterns; a third layer might combine the results of the second layer; and so on.
intelligence, the rights of the software developer may be called into question. In that scenario, the legislature or the court may take a pragmatic approach to create a legal fiction to respond to technological change. A paradigmatic example of such a legal fiction is the work-made-for-hire doctrine in copyright law. Although authors have to be humans—as the U.S. Copyright Office recently made clear following the many interesting questions generated by the so-called "monkey selfie"—corporations are nonhuman, institutional authors that have enjoyed protection under the Copyright Act. Given the ready availability of this legal fiction, some commentators have suggested the use of the work-made-for-hire doctrine to address the complicated relationship between machines and authorial works.

---

203. See generally ETHEM ALPAYDIN, MACHINE LEARNING: THE NEW AI 104-09 (2016) (discussing deep learning); KELLEHER & TIERNEY, supra note 140, at 121-30 (discussing neural networks and deep learning); THIERRY POIBEAU, MACHINE TRANSLATION 181-95 (2017) (discussing deep learning in the machine translation context).


205. See U.S. COPYRIGHT OFFICE, CIRCULAR 1: COPYRIGHT BASICS 1 (2017), https://www.copyright.gov/circs/circ01.pdf ("An original work of authorship is a work that is independently created by a human author and possesses at least some minimal degree of creativity."); U.S. COPYRIGHT OFFICE, COMPENDIUM OF U.S. COPYRIGHT OFFICE PRACTICES ch. 300, at 17 (3d ed. 2017) (listing as an example of works that the U.S. Copyright Office will not register a "photograph taken by a monkey").

206. See Naruto v. Slater, 888 F.3d 418, 420 (9th Cir. 2018) (finding that animals, as nonhumans, "lack[] statutory standing under the Copyright Act"). Although the appellate court and commentators often refer to the animal plaintiff as a monkey, Naruto is actually a Sulawesi crested macaque, which is a black ape. See id.

207. See 17 U.S.C. § 201(b) ("In the case of a work made for hire, the employer or other person for whom the work was prepared is considered the author for purposes of this title . . . .").

208. As Bridy observed:

The work made for hire doctrine is a more fitting framework within which to situate the problem of [artificial intelligence] authorship because it represents an existing mechanism for directly vesting ownership of a copyright in a legal person who is not the author-in-fact of the work in question . . . . With respect to works of [artificial intelligence] authorship, treating the programmer like an employer—as the author-in-law of a work made by another—would avoid the problem of vesting rights in a machine and ascribing to a machine the ability to respond to copyright's incentives.
Meanwhile, others have suggested that the ownership questions should be determined on a case-by-case basis. Some commentators have gone even further to call for the recognition of machines as authors and inventors.

It would also avoid the expedient logic that conflates the author's author (i.e., the programmer) with the actual author (i.e., the generative program).

Bridy, supra note 200, at 26 (footnote omitted); see also Bailey, supra note 129, at 171 ("Firms would . . . be anxious to understand if an artificial intelligence would be treated as an employee under the law or could sign legally binding agreements such as no-compete contracts . . . ."); 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, 707-18 (advancing a work-made-for-hire model to govern works generated by AI systems).

209. As Jani McCutcheon explained:

Depending on the facts, there is sufficient flexibility in the provision to accommodate a range of candidates as the deemed author, and it would be imprudent and unnecessary to nominate only one for all circumstances. Instead, all relevant factors should be considered and balanced against each other. While it may be convenient to have greater consistency and predictability with a 'bright-line' rule, a case by case consideration may be the only possible approach.

Jani McCutcheon, Curing the Authorless Void: Protecting Computer-Generated Works Following IceTV and Phone Directories, 37 Melb. U. L. Rev. 46, 68 (2013); see also Robert C. Denicola, Ex Machina: Copyright Protection for Computer-Generated Works, 69 Rutgers U. L. Rev. 251, 286-87 (2016) ("A computer user who initiates the creation of computer-generated expression should be recognized as the author and copyright owner of the resulting work.").

210. As Ryan Abbott observed in the patent context:

I argue that we should recognise computers as inventors. This will functionally produce more invention because it will incentivise the development of creative computers. That is because allowing computer owners to patent the output of their machines makes those machines more valuable. The constitutional rational[e] for granting patent inventions in the United States is based on an incentive theory. We want patents because of the free-rider problem and because patents are thought to generate additional research and discovery. Even though computers do not care about incentives, people who design computers do. Acknowledging computers as inventors would reward effort upstream of the stage of invention, and it could also promote disclosure and commercialisation of patentable subject matter.

Ryan Abbott, Inventive Machines: Rethinking Invention and Patentability, in INTELLECTUAL PROPERTY AND DIGITAL TRADE, supra note 15, at 113, 117-18; see also Bailey, supra note 129, at 171 (noting that the wrong type of legislation in this area "would damper [the firms'] enthusiasm to build intellectually creative machines whose work would automatically belong to the public domain"). See generally Madeleine de Cock Buning, Artificial Intelligence and the Creative Industry: New Challenges for the EU Paradigm for Art and Technology by Autonomous Creation, in RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE 511 (Woodrow Barfield & Ugo Pagallo eds., 2018) (discussing the complications autonomous creation would pose to intellectual property law); Liza Vertinsky, Thinking Machines and Patent Law, in RESEARCH HANDBOOK ON THE LAW OF ARTIFICIAL INTELLIGENCE, supra, at 489 (discussing the challenges that a thinking machine paradigm of invention would pose to U.S. patent law).
In sum, in the example concerning data produced by the networked sensors in a smart car, complications will arise over who should have control over the generated data. While this subpart has shown that car owners, car manufacturers, and software manufacturers all have strong claims to machine-generated data, one should not overlook the possibility of a fourth, and highly attractive, option: the release of these data into the public domain for all to use. As Drexl rightly noted, before deciding who owns what, we need to first determine whether data should be owned in the first place. Likewise, Reto Hilty warned that asking the question "who 'owns' the data? . . . is already suggestive, because the primary question should be: can (or should) data be 'owned' at all?"

Indeed, even if the answers to these questions are in the affirmative, one does not always have to pick between these three groups of players. One could easily support co-ownership arrangements among some or all of these players, similar to what the joint ownership of copyrighted works already allows. When such

---

211. Who has initial control may not matter as much if the transactions costs are low. As the European Commission noted in its staff working document:

According to the Coase theorem, the effects of an initial allocation of a good may be limited if the good is freely tradable and the transaction costs are sufficiently small. Depending on market forces/bargaining position, it is therefore possible that rights in data would be traded away to the actor(s) who would most benefit from its use. 

Commission Staff Working Document, supra note 10, at 36 (footnote omitted). Nevertheless, the discussion of the "tragedy of the anti-commons" in Part III.A suggests the existence of holdups that prevent effective Coasian bargaining. See Kerber, supra note 5, at 997 ("The problem of patent hold-ups for complex products (that need the simultaneous consent of many patent holders) can also arise in regard to data . . .").

212. See Hugenholtz, supra note 6, at 49 (noting four possibilities of data ownership: "the automobile manufacturer; the car owner; the producer of the sensor equipment; or no one at all").

213. As Drexl observed:

[The question "Who owns the data?" is fundamentally misguided. . . . [I]t skips the prior question of whether there is a need to recognize any ownership. There is no natural law that says that data as an asset, although it may have economic value, has to be owned by anybody. Rather, recognition of any new right should, as is the case in intellectual property in general, be considered a form of government regulation of the market, which is in need of a particular justification. In terms of data ownership, which enables its owner to commercialize data, this justification needs to be an economic one.

Drexl, supra note 17, at 260.

214. Hilty, supra note 144, at 88-89.

215. See id. at 91 (raising the question whether the complex setting involving ownership of data in a traffic app should "entail a kind of 'co-ownership'".)

216. See 17 U.S.C. § 201(a) (2012) ("The authors of a joint work are coowners of copyright in the work.").
co-ownership arrangements occur, new questions will arise over the specific interests each player owns and what specific rights these interests entail. Co-ownership arrangements will also likely create more transaction costs than those involving sole ownership, thereby exacerbating the concern about creating a thicket of data producer’s rights.

3. Duration of Protection

The third set of policy questions pertains to the duration of the proposed data producer’s right. The EU Database Directive offers protection for about fifteen years that can be indefinitely extended based on a “substantial change . . . to the contents of a database, including any substantial change resulting from the accumulation of successive additions, deletions or alterations.”218 The proposed, but eventually abandoned, Database Treaty contained language offering protection for an even longer duration—twenty-five years, indefinitely extendable under similar conditions.219 Based on these prior developments, one may wonder whether the proposed data producer’s right should last for at least fifteen years.

Nevertheless, if the goal of this proposed right is to ensure the opportunity for data producers and their collaborators to analyze the generated data, the protection may not need to last for such a long period of time.220 For instance, the original proposal for the EU Database Directive offered protection for only ten years.221 Indeed,

217. See Drexl, supra note 17, at 277 (“[I]f everybody contributing to the generation of data in a value network is vested with ownership, this allocation could easily run the risk of creating too many property rights, which would block efficient exploitation of big data in particular.”).
219. WIPO, supra note 44, art. 8(1) (Alternative A).
220. See Kerber, supra note 5, at 991 (noting the consensus in Germany and Europe that the proposed data producer’s right “should have a limited (and rather brief) duration of protection, e.g., two or five years (perhaps with the possibility of extension), and should be fully tradable”); Zech, supra note 173, at 469 (“A short term of protection would be appropriate.”); see also Drexl, supra note 17, at 278 (“In an environment where it is key to capture the moment and where being late leads to wrong decisions, asking the question of how long data should be protected will simply miss the needs of this economy.”).
221. As Reichman and Samuelson observed, “The first proposed draft E.C. Directive . . . provided a ten-year period of lead time in which the database maker could recoup his or her investment in a noncopyrightable electronic database while preventing copiers from engaging in for-profit extraction or reutilization of the factual contents, in whole or in substantial part.” Reichman & Samuelson, supra note 43, at 81; see also Nat’l Research Council, supra note 69, at 10 (“If [database] legislation with a fixed term of protection is
Reichman and Uhlir have found the fifteen-year term under the EU Database Directive "completely arbitrary."\(^\text{222}\)

Moreover, with the latest technology, data analysis can be done fairly quickly, not to mention that much of this analysis has to be done in real time.\(^\text{223}\) Even Zech, the leading proponent of the proposed data producer’s right, does not believe that the right should last for too long.\(^\text{224}\) Those European commentators in support of creating such a right also seem to have reached a consensus that the duration of protection could be as short as two to five years, with possible extension.\(^\text{225}\)

Finally, law and economics literature has shown that a market head start could help innovators maintain a healthy competitive edge.\(^\text{226}\) Such a head start is similar to the limited competitive advantage the United States Court of Appeals for the Second Circuit provided to sports leagues and their licensees in *NBA v. Motorola, Inc.*\(^\text{227}\) In that case, the court created a "hot news" doctrine to allow the "producers" of data through professional basketball games to retain limited control

\(^{222}\) Reichman & Uhlir, *supra* note 61, at 817.

\(^{223}\) See Exec. Office of the President, *Interim Progress Report*, *supra* note 23, at 2 ("[D]ata analysis is increasingly conducted in speeds approaching real time."); Drexl, *supra* note 17, at 264 ("To keep up with the speed of this process is key in big data analytics because the users of the results of such analyses will usually have to rely on real-time analyses for decision-making in a constantly changing world.").

\(^{224}\) See Zech, *supra* note 173, at 469 (supporting a "short term of protection").

\(^{225}\) See Kerber, *supra* note 5, at 991 (noting that consensus in the German and European discussion that the new data producer’s right "should have a limited (and rather brief) duration of protection, e.g., two or five years (perhaps with the possibility of extension)").

\(^{226}\) See Reichman & Samuelson, *supra* note 43, at 141 ("[D]atabase makers need a market-preserving period of lead time during which unfair competition law may protect them against ‘cloning’ or ‘partial cloning,’ that is, against the wholesale reproduction of all or a substantial component of database contents."); see also Stephen Breyer, *The Uneasy Case for Copyright: A Study of Copyright in Books, Photocopies, and Computer Programs*, 84 Harv. L. Rev. 281, 299-308 (1970) (arguing that the rewards created by a market head start may provide incentives for authors to create).

\(^{227}\) See 105 F.3d 841, 843 (2d Cir. 1997) (prohibiting copying of "hot news" or time-sensitive materials for competition purposes). As McManis observed:

The concept of providing artificial "lead-time" to minimize harm to a database developer’s market is reflected in the state common law "hot news" misappropriation doctrine. Providing a limited term of such artificial lead time allows database developers to recoup their investment of research and development costs. If free riders can immediately and perfectly copy the database, the original developer will be deprived of the opportunity to develop its market niche.

McManis, *supra* note 42, at 23 (footnotes omitted).
over the in-progress transmission of these data.\textsuperscript{228} Notwithstanding the benefits provided by a head start, how much of a head start data producers should have largely depends on empirical analysis. The head start they need may also depend on the state of technology, with variations from sector to sector.\textsuperscript{229}

4. Scope of Protection

The fourth set of policy questions regards the scope of protection. This question is particularly important because a carefully defined scope will help prevent or address the problem of “overlapping rights”—that is, the existence of multiple, and often competing, layers of rights covering the same subject matter.\textsuperscript{230} As Hugenholtz observed:

[A] film shot with a digital camera would qualify not only as a work protected by copyright, but also as machine-generated (sensor) data subject to a “data producer’s right”. Similarly, the aggregate stock market data in a financial database would be protected both by the sui generis right and the “data producer’s right”, since the data are recorded automatically by the computerized stock exchange.\textsuperscript{231}

\textsuperscript{228} See \textit{NBA}, 105 F.3d at 843 (holding that “a narrow ‘hot-news’ exception does survive preemption [by the federal Copyright Act]” and that the “transmission of ‘real-time’ NBA game scores and information tabulated from television and radio broadcasts of games in progress does not constitute a misappropriation of “hot news””); \textit{see also} \textit{VICTORIA SMITH EKSTRAND, HOT NEWS IN THE AGE OF BIG DATA: A LEGAL HISTORY OF THE HOT NEWS DOCTRINE AND IMPLICATIONS FOR THE DIGITAL AGE 159-201 (2015) (discussing the evolution of hot news doctrine after NBA and its potential future in the age of big data)}.

\textsuperscript{229} As Drexl observed:

[A]location of data ownership is . . . an issue of considerable complexity because of the particularities of the specific sectors. The interests of stakeholders regarding the data collected by the sensors of a car, in which public authorities also have an interest, so as to protect the environment or to increase driving safety, are likely to be different than those in the case of health-related data derived from blood tests of patients for which a patented diagnostic tool is used, which, taken together with similar data from other labs, may help authorities around the globe to fight the spread of infectious diseases.

Drexl, \textit{supra} note 17, at 260; \textit{see also} \textit{Commission Communication, supra} note 2, at 17 (calling for “[s]ector-specific experiments on standards”); \textit{MANYIKA ET AL., supra} note 4, at 8 (“Illustrating differences among different sectors, if we compare the historical productivity of sectors in the United States with the potential of these sectors to capture value from big data (using an index that combines several quantitative metrics), we observe that patterns vary from sector to sector . . .”).

\textsuperscript{230} For discussions of overlapping rights, see generally \textit{ESTELLE DERCLAYE \& MATTHIAS LEISTNER, INTELLECTUAL PROPERTY OVERLAPS: A EUROPEAN PERSPECTIVE} (2011); \textit{OVERLAPPING INTELLECTUAL PROPERTY RIGHTS, supra} note 48; \textit{Mark A. Lemley, Dealing with Overlapping Copyrights on the Internet, 22 U. DAYTON L. REV. 547 (1997)}.

\textsuperscript{231} Hugenholtz, \textit{supra} note 6, at 62.
In its staff working document, the European Commission made a distinction between semantic and syntactic information:

An ebook or a photographic image . . . has a semantic level which is the expression of ideas or the presentation of objects or persons. Copyright covers this level of information. The data file of such an ebook or image, however, is merely a representation of signs encoding such information usually requiring tools to present the information.\(^{232}\)

While the distinction between semantic and syntactic information is thought-provoking—and could help address the much-needed divide between copyright and the proposed data producer’s right—it is unclear if lawmakers, policy makers, litigants, and law enforcement personnel could easily make that distinction. As Hugenholtz explained:

\[\text{[A]ny copy of the film’s digital file (the syntactic data) would by necessity also reproduce the copyright protected work (the semantic layer). Thus, the new data right could be invoked against any digital copying (or streaming) of the digitized copyright work. For the same reason, the new right would broadly overlap with the database right, even if its scope were confined to the syntactic layer. The phonographic right discussed above illustrates this point. Whereas its subject matter, like the proposed “data producer’s right”, is limited to the recorded signal (i.e. syntactic audio data), its scope extends into the semantic realm. Reproducing a CD recording of a musical performance will, by necessity, result in the reproduction of the underlying musical work and performance.}\(^{233}\)

---

\(^{232}\) Commission Staff Working Document, supra note 10, at 34 n.156; see also Drexl, supra note 17, at 263 (“[A] distinction can be made in terms of semiotics between the different levels of information. For data protection, the distinction between the syntactic and the semantic level is key.” (footnote omitted)). As Zech explained:

\[\text{A distinction can be made on the level of meaning (semantic information), such is the case with personal data defined as information relating to a person . . . . Know-how is also considered to be semantic information, when defined by its meaning.}\]

\[\text{The term “data” can also be defined on the level of signs (syntactic information), regardless of its meaning. As a legal object, sequences of “zeros” and “ones” would be protected, either as a file or as a data stream. This distinction is comparable to other syntactic representations of information, such as text (defined by its symbols rather than meaning).}\]

Zech, supra note 173, at 462-63 (footnote omitted).

\(^{233}\) Hugenholtz, supra note 6, at 63-64; see also Scassa, supra note 1, at 4 (“[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.”).
Consider, for example, the use of multiple camera drones to shoot a telecast of a professional basketball game, similar to the one involved in NBA. As a work consisting of sounds and images that are “being made simultaneously with its transmission,” the telecast will constitute a copyrighted work, with the league or its designated licensing arm, NBA Properties, Inc., being the work’s copyright holder. Using the definition provided by the European Commission, the telecast would fall squarely within the category of semantic information. By contrast, the statistics about the professional basketball players involved—such as minutes played, points scored, rebounds, assists, steals, blocks, and turnovers—will be classified as syntactic information. As such, the data generated by the camera drones will be protected by the proposed data producer’s right.

Yet when all of the syntactic information has been inputted into a computer that stores three-dimensional image files relating to professional basketball players, similar to those found in the video game NBA Live, the computer-generated telecast will bear a strong resemblance to the copyrighted telecast, blurring the distinction between semantic and syntactic information. In this instance, the interplay of these two types of information will certainly worry the NBA. If licensing arrangements are to be developed, the different types of right involved will also raise considerable complications.

At some point, policy makers and commentators will have to recognize the inability to use a separate binary to separate semantic and syntactic information. Rather, all the information will fall on a spectrum, in which syntactic information, as numbers and symbols, could be combined to form meaning and be transformed into semantic information. This spectrum is similar to the one involving unstructured, semi-structured, and structured data—with semantic information on the end with unstructured data and syntactic information on the other end.

234. 105 F.3d 841 (2d Cir. 1997).
235. See 17 U.S.C. § 101 (2012) (“A work consisting of sounds, images, or both, that are being transmitted, is ‘fixed’ for purposes of this title if a fixation of the work is being made simultaneously with its transmission.”).
236. See NBA, 105 F.3d at 841 (listing NBA Properties, Inc. as a plaintiff).
237. See HOLMES, supra note 1, at 5-6 (noting the oft-used taxonomy of structured, semi-structured, and unstructured data). Semi-structured data are defined as “[d]ata that do not conform to fixed fields but contain tags and other markers to separate data elements.” MANYIKA ET AL., supra note 4, at 33. Examples of such data “include XML or HTML-tagged text.” Id.
To complicate matters even further, there is no hard-and-fast rule for determining whether it is more desirable to protect semantic or syntactic information. Using the example of a pothole located by car sensors, Drexl explained why “it would be better to avoid protecting the semantic information the sensors of a car collect,” even though intellectual property rights generally protect this type of information.\(^{238}\) As he observed:

> [T]he question of whether the law should protect the semantic or the syntactic information, or even only the integrity of the digital file, will depend on the circumstances. This analysis would seem to argue for context-specific regulation. Even a general regime on the protection of industrial data would thus appear problematic since, in some instances, protecting semantic information such as in the case of trade secrets seems the right approach, while protection of data collected through sensors in the public sphere should probably not be extended to the meaning these data are able to convey.\(^{239}\)

5. Limitations and Exceptions

The last set of policy questions involves the creation of limitations and exceptions—or, more positively, “the modalities of access.”\(^{240}\) As Timo Minssen and Justin Pierce lamented: “Where data are not owned or licensed, the user will need to rely on an exception to [intellectual property right] infringement to use data. This has given rise to fierce controversies between data ‘owners,’ data researchers, and entities that provide enabling technologies, large research infrastructures, and standardization platforms.”\(^{241}\) Although data are often referred to as “new oil,”\(^{242}\) one cannot ignore the fact that the latter is a finite resource.\(^{243}\) Because of this crucial distinction, it could be quite

\(^{238}\) Drexl, supra note 17, at 263.

\(^{239}\) Id.


\(^{241}\) Minssen & Pierce, supra note 177, at 311; see also REICHMAN ET AL., supra note 8, at 367 (“[N]ational and international intellectual property laws . . . remain hostile to the needs of digitally integrated scientific research, and especially to publicly funded and public interest research endeavors.”).

\(^{242}\) See sources cited supra note 1.

\(^{243}\) See HOLMES, supra note 1, at 20 (“[U]nlike oil, data appears not to be a finite resource.”); Scassa, supra note 1, at 1 (“[M]any are quick to point out that data are an infinitely renewable resource.”); Lauren Henry Scholz, Big Data Is Not Big Oil: The Role of Analogy in the Law of New Technologies (Fla. State Univ. Pub. Law & Legal Theory, Paper No. 895, 2019), https://ssrn.com/abstract=3252543 (challenging the data-as-oil analogy); see also ORG.
beneficial for data to be used simultaneously by multiple parties or reused by others after their earlier use.\textsuperscript{244}

One potential exception to the data producer’s right that would facilitate data use and reuse is the exception for text and data mining (TDM). A similar exception is currently being explored as part of the copyright law reform in Australia,\textsuperscript{245} the European Union,\textsuperscript{246} Singapore,\textsuperscript{247} and other parts of the world.\textsuperscript{248} In the United Kingdom, such an exception can now be found in section 29A of the Copyright, Designs and Patents Act 1988, which allows “[t]he making of a copy of a work by a person who has lawful access to the work” in order to “carry out a computational analysis of anything recorded in the work for the sole purpose of research for a non-commercial purpose.”\textsuperscript{249}

---

\textsuperscript{244} See Scassa, supra note 1, at 1 (“The same data are... capable of being used by multiple actors and for multiple purposes.”).


\textsuperscript{246} See Drexl, supra note 17, at 276 (noting the “proposal of the Commission to introduce an unwaviable exception to copyright protection for carrying out text and data mining for the purpose of scientific research”).

\textsuperscript{247} See Ministry of Law, Public Consultation on Proposed Changes to Singapore’s Copyright Regime 34-35 (2016) (seeking consultation for Proposal 9 on TDM).

\textsuperscript{248} See Christophe Geiger et al., Crafting a Text and Data Mining Exception for Machine Learning and Big Data in the Digital Single Market, in Intellectual Property and Digital Trade, supra note 15, at 95, 97 (“Text and data mining... serves as an essential tool to navigate the endless sea of online information in search of this invaluable treasure that big data might hold for the European economy.”); Xavier Seuba et al., Introduction to Intellectual Property and Digital Trade, supra note 15, at 9, 15 (“Text and data mining... has been a fundamental technique to make machine learning—and artificial intelligence autonomous decision-making and creativity—possible by copying or crawling massive datasets.”). See generally Matthew Sag, The New Legal Landscape for Text Mining and Machine Learning, 66 J. Copyright Soc’y U.S.A. (forthcoming 2019) (discussing the changing U.S. legal landscape for TDM research).

\textsuperscript{249} Copyright, Designs and Patents Act 1988, c. 48, § 29A(1) (Eng.). As Christophe Geiger, Giancarlo Frosio, and Oleksandr Bulayenko explained, this exception can help avoid copyright infringement in many situations:

[A]ny reproductions resulting in the creation of a copy of a protected work along the chain of TDM activities might trigger copyright infringement. In this respect, pre-processing to standardize materials into machine-readable formats might trigger infringement of the right of reproduction. Likewise, the uploading of the pre-processed material on a platform—which might or might not occur, depending on whether the TDM technique makes use of TDM software crawling data to be analysed directly from the source—might also violate the right of reproduction. Mining (the stage of the TDM process where data are finally extracted) can also infringe upon the right of reproduction, depending on the mining software deployed and the character of the extraction. For example, there are extraction techniques that
Other exceptions that could be of great public benefit are those concerning government access,\textsuperscript{250} scientific research,\textsuperscript{251} data portability,\textsuperscript{252} would reproduce parts of the work so minimal as to fall below the threshold of copyright infringement.

Again, TDM might involve the reproduction, translation, adaptation, arrangement, and any other alteration of a database protected by copyright, which means the original selection and arrangement of the database's content. For example, pre-processing for extraction might cleanse from a database portions and data that are irrelevant for data analysis. In this respect, pre-processing might violate both the right of reproduction and the right to make adaptations and arrangements.

Geiger et al., \textit{supra} note 248, at 98-99 (footnotes omitted).

250. \textit{See Commission Staff Working Document, supra} note 10, at 36 ("[P]ublic sector bodies may ... have a legitimate interest in obtaining access to certain data. This has relevance for the provision of statistical information, urban planning, environmental protection, civil protection, etc. In most situations, public sector bodies would need aggregate information only."); Kelleher & Tierney, \textit{supra} note 140, at 223-24 (discussing the use of data by city authorities to monitor and manage traffic); Drexl, \textit{supra} note 17, at 262 ("[T]he geographical location of the car at a given moment can ... inform the public authorities about the volume of use and traffic conditions of roads at a given time."); Farkas, \textit{supra} note 10, at 6 (noting that the data communicated by networked cars "can be used to ... improve road conditions ... and to plan road reconstructions"); Reichman & Samuelson, \textit{supra} note 43, at 116-17 ("Much of the data that fuels some scientific disciplines is collected by governments to foster public safety (for example, providing timely warnings of potential disasters, such as floods, tornadoes and the like).""). As Drexl explained:

[A]ccess to data is justifiable where public entities seek access for the fulfilment of tasks in the public interest. In the light of the large benefits deriving from big data analytics, which could help optimise public policies and decisions of the state in many regards, this sub-category for which access regimes could be implemented seems most important. Such regimes could be implemented at the different levels of government through sector-specific regulation. Sector-specific regulation appears as the road to take, since the security interests of the state will most likely need different rules than the prevention of infectious diseases, the protection of the environment or the functioning of smart cities or traffic control systems.

Drexl, \textit{supra} note 17, at 289 (footnote omitted).

251. \textit{See Commission Staff Working Document, supra} note 10, at 36 ("[T]he Commission's policy on open science and open access, an exception ensuring access to relevant privately-held data could be considered for scientists performing research entirely or predominantly funded by public resources.").

252. \textit{See Commission Communication, supra} note 2, at 15 (discussing the portability of nonpersonal data); \textit{Commission Staff Working Document, supra} note 10, at 46-49 (same); Drexl, \textit{supra} note 17, at 286 ("Since this rule on data portability constitutes a most suitable form of pro-competitive regulation, there is no reason why the right to data portability should be limited to personal data."); Kerber, \textit{supra} note 5, at 997 ("[S]upporting portability, interoperability and standardization in regard to data is seen as pivotal policy measures for improving the governance of data in the digital economy."); \textit{see also} Council Regulation 2016/679, \textit{supra} note 26, art. 20 (introducing the right to data portability). Nevertheless, the Commission noted in its communication document:

[R]egarding non-personal data, there are at present no obligations to guarantee even a minimum level of data portability, even for widely used online services such as cloud hosting providers. This is partly because the requirements for implementing
platform interoperability,\(^\text{253}\) the nonprotection of public sector information,\(^\text{254}\) and the need to address anticompetitive concerns.\(^\text{255}\)

To promote competition, the European Commission suggested the use of fair, reasonable, and nondiscriminatory (FRAND) licenses to facilitate the utilization and dissemination of machine-generated data.\(^\text{256}\) As the Commission stated in its communication document, "[F]ramework potentially based on certain key principles, such as [FRAND] terms, could be developed for data holders . . . to provide

data portability can be technically demanding and costly, as different providers of the same services may store data differently.

Commission Communication, supra note 2, at 15.

253. See Commission Communication, supra note 2, at 16 (discussing the interoperability of nonpersonal data); Mayer-Schönberger & Cukier, supra note 1, at 183 ("We should enable data transactions, such as through licensing and interoperability."); Drexl, supra note 17, at 292 ("The functioning of the data economy will also depend on the interoperability of digital formats and the tools of data collecting and processing."). See generally Michal S. Gal & Daniel L. Rubinfeld, Data Standardization, 94 NYU L. Rev. (forthcoming 2019) (providing an excellent discussion of data standardization).

254. See Commission Staff Working Document, supra note 10, at 22 (subjecting the re-use of data held by the public sector to rules under the Directive 2003/98/EC); Org. for Econ. Co-Operation & Dev., supra note 4, at 403-38 (examining the benefits and challenges of opening access to data from the public sector); Drexl, supra, note 17, at 262 ("[S]tates started to realise that it is becoming increasingly important to grant private businesses access to publicly held data . . . for commercial re-use in order to promote new commercial information services."); Scassa, supra note 1, at 2-3 ("[A]ny data ownership rights must similarly include exceptions that are appropriate for the public interest."); see also Nat'l Research Council, supra note 69, at 100 ("Protection should not be extended to database collected or maintained by the government." (emphasis omitted)); Ruth L. Okediji, Government as Owner of Intellectual Property? Considerations for Public Welfare in the Era of Big Data, 18 Vand. J. Ent. & Tech. L. 331, 356-60 (2016) (advancing three approaches to "creatively address[] the terms of access to knowledge goods created from big data").

255. For discussions of the interplay of the proposed data producer's right and competition law, see generally Mayer-Schönberger & Cukier, supra note 1, at 182-84; Drexl, supra note 17, at 280-85; Timo Minssen & Jens Schovsbo, Big Data in the Health and Life Sciences: What Are the Challenges for European Competition Law and Where Can They Be Found?, in Intellectual Property and Digital Trade, supra note 15, at 121.

256. See Commission Communication, supra note 2, at 13; see also Commission Staff Working Document, supra note 10, at 39 ("The degree of openness can range from full openness (the licensor is required to allow free access to data) to a set of intermediate options. Licensing conditions should be fair, reasonable and non-discriminatory between different licensees."). As Wolfgang Kerber observed:

The problem of patent hold-ups for complex products (that need the simultaneous consent of many patent holders) can . . . arise in regard to data, if the provider of a certain product or service needs access to complementary data. It might then be necessary to consider access solutions similar to standard-essential patents (with FRAND-conditions).

Kerber, supra note 5, at 997.
access to the data they hold against remuneration after anonymisation." 257

Finally, to ensure the effectiveness and vitality of these limitations and exceptions, some jurisdictions may consider the adoption of "contract override" legislation to prohibit the use of private contracts to circumvent the safeguards provided by these limitations and exceptions. 258 In August 2014, the United Kingdom adopted the Copyright and Rights in Performances (Quotation and Parody) Regulations 2014, 259 which rendered unenforceable contractual terms that purport to prevent or restrict the newly created permitted acts, such as those relating to quotation, caricature, parody, and pastiche. 260

To the extent that the limitations and exceptions are highly beneficial to the public, a contract override would prevent parties with weak bargaining positions from being forced to contract away those important flexibilities and safeguards. 261

B. Exogenous Complications

The previous subpart focused on the proposed data producer's right as a new form of protection within the intellectual property system. This subpart turns to four related areas in which the proposed right would create additional complications. The first set of complications is still within the intellectual property system, while the other three sets

258. See Drexl, supra note 17, at 276 ("Access can also be guaranteed by special legislation on access that takes precedence over contractual restrictions."); id. at 291 ("In principle, the legislature could ... promote access through un-waivable exceptions and limitation as part of a comprehensive legislation of data ownership."); see also J.H. Reichman & Jonathan A. Franklin, Privately Legislated Intellectual Property Rights: Reconciling Freedom of Contract with Public Good Uses of Information, 147 U. PA. L. REV. 875, 929-38 (1999) (advancing a "doctrine of public-interest unconscionability"). For discussions of contractual override or issues relating to "contracting out," see generally Austl. Law Reform Comm'n, Copyright and the Digital Economy: Discussion Paper 353-77 (2013); Austl. Law Reform Comm'n, supra note 245, at 435-57; Reichman et al., supra note 8, at 349-50.
259. The Copyright and Rights in Performances (Quotation and Parody) Regulations 2014, SI 2014/2356 (Eng.).
260. See Copyright, Designs and Patents Act 1988, c. 48, § 30A(2) (Eng.) ("To the extent that a term of a contract purports to prevent or restrict the doing of any act which, by virtue of this section, would not infringe copyright, that term is unenforceable.").
261. See Commission Communication, supra note 2, at 12 ("The unequal bargaining power of companies and private individuals should be taken into account. Lock-in situations, especially for [small and medium-sized enterprises] and startups and private individuals, should be avoided."). Notably, article 15 of the EU Database Directive states that "[a]ny contractual provision contrary to Articles 6(1) and 8 shall be null and void." Council Directive 96/9, supra note 31, art. 15. Both provisions cover the rights and obligations of lawful users. See id. arts. 6(1), 8.
of complications lie outside. A deeper understanding of these four sets of complications will be highly significant in the development of a sound governance regime to address the emerging needs of our fast-growing data-driven economy.

1. Intellectual Property

The first set of complications concerns the protection of other forms of intellectual property rights. As noted earlier, the proposed data producer’s right could generate the problem of “overlapping rights.”

For example, the new protection for machine-generated data could create complications with the preexisting protection for not only databases but also data themselves. While the EU Database Directive continues to use copyright to protect databases that constitute intellectual creations, it also offers *sui generis* protection to databases that are ineligible for copyright protection. The proposed data producer’s right would add a new layer of protection for uncopyrightable machine-generated data to these two layers of rights even when the investments involved have been made for “the creation of materials which make up the contents of a database,” as

---

262. See discussion supra Part IV.A.4.

263. See Council Directive 96/9, supra note 31, art. 3(1) (“[D]atabases which, by reason of the selection or arrangement of their contents, constitute the author’s own intellectual creation shall be protected as such by copyright.”).

264. See id. arts. 7-11 (providing coverage for the *sui generis* database right).


[A] *sui generis* database right only subsists if “there has been qualitatively and/or quantitatively a *substantial investment* in either the *obtaining, verification or presentation of the contents*. The [Court of Justice of the European Union (CJEU)] has interpreted these requirements in a very restrictive way. It clarified that the investment has “to refer to the resources used to seek out existing independent materials and collect them in the database, and not to the resources used for the creation as such of independent material[s].” The CJEU explained this with the objective of the Directive to create incentives for the making of databases and not for the creation of the data that goes into the database. Hence, a distinction is to be made between the “creation” of the materials contained in the database and the “obtaining” of these materials. This leads to the conclusion that the creation of smart products with sensors that collect data should not be considered for the assessment of whether the investment in the database was “substantial”. The same applies to big data analyses. These may well require substantial investment. However, such analyses only lead to the creation of new data in the form of knowledge, which may
opposed to the "obtaining, verification or presentation of the[se] contents."²⁶⁶

If one focuses on data instead, the proposed data producer’s right might create further complications with preexisting laws covering trade secrets²⁶⁷ and data exclusivities.²⁶⁸ After all, these alternative forms of intellectual property rights have already been used to protect data. Indeed, the European Commission’s staff working document suggests that the proposed data producer’s right could take the form of a defensive right.²⁶⁹ Such a right would likely be quite similar to the protection currently offered under trade secrecy or unfair competition law.²⁷⁰

When overlaps arise, the proposed right could lead to regime clashes and overprotection.²⁷¹ Such clashes and overlaps, in turn, would undermine the protection offered in preexisting intellectual property regimes, especially in relation to their limitations, safeguards, and flexibilities.²⁷² If these overlaps are to be avoided, commentators have suggested the need for preemption tools.²⁷³ As
Hugenholtz observed, "The only way to prevent the data right from becoming an all-encompassing ‘super-[intellectual property] right’ would be to categorically exclude all data that (possibly) represent subject matter protected under traditional [intellectual property] regimes: not only copyright, the database right and neighboring rights, but also design right and perhaps patent law."274

2. Privacy

The second area that the proposed data producer’s right may spill into pertains to the protection of personal data, which is in the field of privacy or data protection law. With the implementation of the EU General Data Protection Regulation (GDPR)275 in May 2018, and the creation of considerable ramifications for individuals and businesses in the European Union and abroad, the complications of privacy law have caught the attention of policy makers, commentators, private businesses, and individuals.

To be sure, the GDPR and other privacy laws cover personal data, while the proposed data producer’s right focuses on nonpersonal, anonymized machine-generated data.276 Nevertheless, the increased sophistication of data analytics has raised questions about the effectiveness of data anonymization. A growing volume of research has already shown the possibility of reconstructing seemingly anonymized data to reveal the identity of data subjects.277


274. Hugenholtz, supra note 6, at 64.
276. As the European Commission declared in its communication document:
Where personal data are concerned, the individual will retain his right to withdraw his [or her] consent at any time after authorising the use. Personal data would need to be rendered anonymous in such a manner that the individual is not or no longer identifiable, before its further use may be authorised by the other party. Indeed, the GDPR continues to apply to any personal data (whether machine generated or otherwise) until that data has been anonymised.

Commission Communication, supra note 2, at 13.

277. See MAYER-SCHÖNBERGER & CUKIER, supra note 1, at 154 (“[B]ig data, with its increase in the quantity and variety of information, facilitates re-identification.”); Cate, supra note 15, at 9 (“[I]n a world of big data . . . with sufficient, interconnected data, even de-identified or anonymized data may be rendered personally identifiable.”); Ohm & Peppet, supra note 15, at 45 (“[B]ecause of big data advances in data analytics, we may soon learn that ‘everything reveals everything.’”); Peppet, supra note 12, at 118-23 (discussing how sensor fusion and big data analytics may mean that everything reveals everything).
Moreover, seemingly anonymized data may be highly personal. Consider, for instance, the data about car locations appearing on navigation apps. As Hilty explained:

The colour green [on these apps] is used for flowing traffic, orange for slow-moving traffic, and red to signal a traffic jam. This information is not collected based on hundreds of helicopters or drones flying over the country, sending pictures to traffic control centres. Instead, this information is generated by correlative movements of the mobile telephones of the car drivers passing through the same positions; Apple, Android, and all other applications are based on similar technologies. It goes without saying that individual drivers could be identified; their data could for instance be connected with data produced by the car itself, or advertisements could be sent to them, for example for nearby restaurants when there is a traffic jam.278

In sum, given the potential, and arguably inevitable, overlap between personal and nonpersonal data, policy makers should develop a deeper understanding of the interplay between the protection for personal or personally identifying data and the protection for nonpersonal, anonymized data that the proposed data producer’s right seeks to offer. If policy makers are unsure how to protect the former, they should pause to think more deeply about the implications of the proposed data producer’s right for privacy protection.

3. Trade

The third area that the proposed data producer’s right may spill into relates to the ongoing discussion of digital trade and cross-border data flows at the World Trade Organization (WTO) and other international and regional fora.279 Many of the bilateral, regional, and plurilateral trade agreements that are under negotiation or that have recently been completed also include dedicated chapters to promote the free movement of data.280 To some extent, the electronic commerce chapters included in these agreements “represent... a distinct attempt

---

278. Hilty, supra note 144, at 91.
280. See Burri, supra note 279, at 99-125 (discussing developments relating to U.S. free trade agreements and the TPP).
to compensate for the lack of progress in the WTO and remedy the ensuing uncertainties.\footnote{281}

Included in the TPP electronic commerce chapter are a wide variety of issues, such as customs duties, nondiscriminatory treatment of digital products, domestic electronic transactions, electronic authentication and signatures, online consumer protection, personal information protection, paperless trading, access to and use of the Internet for electronic commerce, cross-border transfer of information by electronic means, Internet interconnection charge sharing, location of computing facilities, unsolicited commercial electronic messages, cooperation on cybersecurity matters, and the source code of computer software.\footnote{282}

As Mira Burri explained, these issues emerge in international and regional fora in part because “data needs to cross borders for a thriving data economy; [but] at the same time, states do exercise jurisdiction within their borders as a rule of public international law.”\footnote{283} In her view, trade rules affect data in three ways:

(i) . . . they regulate the cross-border flow of data by regulating trade in goods and services as well as the protection of intellectual property; (ii) . . . they may install certain beyond-the-border rules that demand changes in domestic regulation—for example, with regard to intermediaries’ liability; and (iii) . . . trade law can limit the policy space that regulators have at home.\footnote{284}

Among the restrictions commentators have identified in this area are “requirements for data localisation, mandates for local processing of data, . . . requiring government approval for data transfers,” as well as measures “that are discriminatory in their treatment of foreign data suppliers.”\footnote{285}

In the trade arena, a widely scrutinized development concerns the negotiation of the TPP electronic commerce chapter.\footnote{286} Considered by many as the most comprehensive electronic commerce chapter among all bilateral, plurilateral, and regional trade agreements,\footnote{287}  

\footnotesize
\begin{enumerate}
  \item \textit{Id.} at 101.
  \item TPP Agreement, \textit{supra} note 169, ch. 14.
  \item Burri, \textit{supra} note 279, at 68.
  \item \textit{Id.}
  \item Maskus, \textit{supra} note 129, at 26.
  \item TPP Agreement, \textit{supra} note 169, ch. 14.
  \item See Burri, \textit{supra} note 279, at 113 (“The TPP chapter on e-commerce is clearly the most comprehensive of all [free trade agreements] so far. It comprises eighteen articles and includes new features that in effect signal an expansion of the U.S. template for digital trade.” (footnote omitted)).
\end{enumerate}
this chapter has now been incorporated as part of the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP). Covering the "Cross-Border Transfer of Information by Electronic Means," article 14.11 of the CPTPP specifically provides: "Each Party shall allow the cross-border transfer of information by electronic means, including personal information, when this activity is for the conduct of the business of a covered person." Titled "Location of Computing Facilities," article 14.13 further provides signatory parties with discretion to "its own regulatory requirements regarding the use of computing facilities, including requirements that seek to ensure the security and confidentiality of communications." Nevertheless, the CPTPP electronic commerce chapter prohibits the application of regulatory measures "in a manner which would constitute a means of arbitrary or unjustifiable discrimination or a disguised restriction on trade.

4. Investment

The final area that the proposed data producer's right may spill into involves the ongoing effort to address intellectual property disputes via the investor-state dispute settlement mechanism. As I noted in an earlier article, the growing use of this mechanism has sparked a transformation similar to what we experienced three


289. TPP Agreement, supra note 169, art. 14.11.2.
291. Id. art. 14.13.3(a).
decades ago when the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) was established to marry intellectual property to trade.

At the time of writing, there is no indication that data producers have plans to follow the leads of Philip Morris, Eli Lilly, and the Japanese Bridgestone Group in using investor-state dispute settlement to protect machine-generated data. Nevertheless, because “most international investment agreements define ‘investment’ broadly to cover all forms of ‘intellectual property rights,”’ these relevant agreements will cover the proposed data producer’s right once that right has been created.

Yu, Investment-Related Aspects, supra note 292, at 831 (footnotes omitted).

295. Philip Morris Brands Sàrl v. Oriental Republic of Uru., ICSID Case No. ARB/10/7, Award (July 8, 2016) (using the investor-state dispute settlement mechanism in the bilateral agreement between Switzerland and Uruguay to challenge tobacco control measures in Uruguay); Philip Morris Asia Ltd. v. Commonwealth of Austl., PCA Case No. 2012-12, Award on Jurisdiction and Admissibility (Dec. 17, 2015) (using the investor-state dispute settlement mechanism in the bilateral agreement between Australia and Hong Kong to challenge tobacco control measures in Australia).

296. Eli Lilly & Co. v. Gov’t of Can., ICSID Case No. UNCT/14/2, Final Award (Mar. 16, 2017) (utilizing Chapter Eleven of the North American Free Trade Agreement to seek compensation for the Canadian courts’ invalidation of its patents on two hyperactivity drugs).

297. Bridgestone Licensing Servs., Inc. v. Republic of Pan., ICSID Case No. ARB/16/34, Request for Arbitration (Oct. 7, 2016) (using the investor-state dispute settlement mechanism in the bilateral agreement between Panama and the United States to challenge the damage award granted by the Supreme Court of Panama in relation to the investor’s action in opposing a trademark registration).

298. Yu, Investment-Related Aspects, supra note 292, at 861. But see TPP Agreement, supra note 169, art. 18.1 (defining “intellectual property” to cover “all categories of intellectual property that are the subject of Sections 1 through 7 of Part II of the TRIPS Agreement”).
In fact, if the proposed right is to be created using the property model, data producers will have rather compelling arguments about their ability to invoke the investor-state dispute settlement mechanism to address disputes involving this new right. After all, a key attraction of this mechanism is to offer "[p]rotection against uncompensated expropriation of property."299 As the Office of the U.S. Trade Representative declared in its fact sheet on investor-state dispute settlement, the mechanism provides "[a]n assurance that the property of investors will not be seized by the government without the payment of just compensation."300

When disputes over ownership of machine-generated data are tied to the right to private property in human rights law, property-based arguments will also find support in international and regional human rights instruments.301 In Anheuser-Busch Inc. v. Portugal,302 the European Court of Human Rights extended the protection of "the peaceful enjoyment of ... possessions" in Protocol No. 1 to the European Convention of Human Rights303 to cover both registered trademarks and trademark applications of a multinational corporation. The right to property provision in the Charter of Fundamental Rights of the European Union also explicitly covers intellectual property.304

---


301. See Yu, supra note 197, at 85-95 (discussing the growing use of the right to property to provide an alternative human rights basis for intellectual property rights and addressing the related concerns).


303. See Protocol to the Convention for the Protection of Human Rights and Fundamental Freedoms art. 1, Mar. 20, 1952, 213 U.N.T.S. 262 ("Every natural or legal person is entitled to the peaceful enjoyment of his possessions. No one shall be deprived of his possessions except in the public interest and subject to the conditions provided for by law and by the general principles of international law.").

304. See Charter of Fundamental Rights of the European Union, art. 17(2), 2000 O.J. (C 364) 1, 12 ("Intellectual property shall be protected.").
C. Summary

The proposal for a new data producer’s right would have complications both within and outside the area of machine-generated data. While the intellectual property system has been rapidly expanding, this Part has shown how the proposed right would create complications that deserve serious policy and academic attention. Given that neither the past lessons nor the present needs support the proposed data producer’s right, the legal and policy challenges this proposal would generate suggest that we should not hastily introduce this proposal without undertaking more extensive study and careful impact assessment. Introducing a data producer’s right could not only harm individuals and businesses but could also affect the rights protected in other areas of intellectual property law as well as in the areas of privacy, trade, and investment.

V. Preliminary Policy Recommendations

This Article has shown that neither past lessons nor present needs support the proposed data producer’s right.305 Worse still, this proposal would raise complications with the future development of a sound and holistic data governance regime.306 The logical conclusion, therefore, is to reject the proposal, similar to how Congress rejected the numerous proposals for *sui generis* database protection in the late 1990s and early 2000s.307 This Article does not stop there, however. It goes further to advance four courses of action that could be instrumental in developing a holistic and effective data governance regime. Such a regime is particularly important considering the current fragmentary legal and regulatory environment in this area.308

First, and most obviously, policy makers should reject the introduction of a data producer’s right for nonpersonal, anonymized machine-generated data, similar to the one advanced by the European Commission.309 If enough empirical evidence suggests that some data

305. See discussion supra Parts II-III.
306. See discussion supra Part IV.
307. See discussion supra Part IIA.
308. See M. Lynne Markus, *Obstacles on the Road to Corporate Data Responsibility*, *in* BIG DATA IS NOT A MONOLITH, supra note 7, at 143, 156 ("[T]he legal and regulatory environment of data protection is fragmented, creating a patchwork of rules with overlaps, conflicts, and gaps.").
309. See Commission Communication, supra note 2, at 13 (suggesting the creation of a new data producer’s right as a possibility for addressing the issue of access to machine-generated data).
producer's right could be, on balance, socially expedient, policy makers should carefully study this evidence to explore whether a limited right is justified—and if so, in what form, in which sector(s), and under what conditions. While there is at present insufficient evidence to support such a limited right, especially one under the property model, this Article does not take the position that such a right can never be justified in any form, in any sector, or under any condition. Whether this right can be justified depends on empirical evidence and impact assessment. The more narrowly the right is defined, the better policy makers will be at scrutinizing the need for such a right.

Second, as Part IV.B has shown, data governance covers many different areas of law and policy. While data governance in the areas of privacy and intellectual property has been widely studied, it is increasingly important to examine data protection in the areas of both trade and investment—areas that are understudied in legal literature.\footnote{310. See Burri, supra note 279, at 76 ("[M]uch of the debate within and outside the WTO, as well as the literature devoted to digital trade, have focused on trade in services and its regulation.").} Although Part IV.B does not further explore the interplay of the proposed data producer's right and competition law\footnote{311. Part IV.B focuses on privacy, trade, and investment, in part because rights in machine-generated data could emerge in these areas. Competition law, by contrast, serves mostly as a countervailing regime. See Drexl, supra note 17, at 261 (noting the "reliance on competition law as a countervailing legal regime").}—a critical area relating to data governance\footnote{312. As Minssen and Pierce observed: While there are issues to be resolved between Big Data and [intellectual property rights], there is a growing awareness of the importance of data and specifically Big Data by market authorities. Antitrust agencies, those in the United States and competition agencies in Europe, are taking note of Big Data, and there is an increasing trend to examine closely the collection, use, and access of Big Data for anticompetitive effects. Minssen & Pierce, supra note 177, at 320; see also Zech, supra note 173, at 461 (noting that competition law is an "important area[] of law for a data economy").}—commentators have advanced notable work in this area.\footnote{313. See Mayer-Schönberger & Cukier, supra note 1, at 182-84 (discussing the use of competition law to govern the data barons); Drexl, supra note 17, at 280-85 (discussing the application of EU competition law to address refusals to grant access to data); Minssen & Schovsbo, supra note 255 (discussing the challenges big data in health and life sciences have posed to EU competition law).} If policy makers are to develop a sound and holistic data governance regime, they will have to break out of their conceptual silos to undertake a holistic examination of issues relating to data governance. Piecemeal examinations, such as those relating to
the data producer's right in the intellectual property context, are unlikely to suffice.

Third, if policy makers do take the position that machine-generated data should not be protected, they need to actively push for prohibitive treaty language at the global level. International intellectual property agreements are filled with language prohibiting the protection of raw data. Article 9.2 of the TRIPS Agreement makes clear that copyright protection shall not extend to "ideas, procedures, methods of operation or mathematical concepts." Article 10.2 further stipulates that the protection of compilations of data or other material "shall not extend to the data or material itself." Likewise, the WIPO Copyright Treaty includes a specific provision on databases:

Compilations of data or other material, in any form, which by reason of the selection or arrangement of their contents constitute intellectual creations, are protected as such. This protection does not extend to the data or the material itself and is without prejudice to any copyright subsisting in the data or material contained in the compilation.

As Part IV.B.3 has shown, the electronic commerce chapters in recent bilateral, regional, and plurilateral trade agreements have focused primarily on issues such as protection of personal information, cross-border transfer of information by electronic means, and location of computing facilities. It is increasingly necessary that policy makers more seriously explore whether these chapters should also include language concerning protection or nonprotection of machine-generated data.

Finally, policy makers and commentators should closely study the five sets of policy questions identified in Part IV.A. Although these questions are difficult to answer, efforts to engage with these questions could be highly valuable. Such engagement could help foster academic debate, promote future legal and policy development, and advance international norm-setting efforts. Indeed, given the similarities between the criticisms of *sui generis* database protection in the 1990s and early 2000s and today's criticisms of the EU proposal for a data

---

314. TRIPS Agreement, *supra* note 293, art. 9.2.
315. *Id.* art. 10.2.
316. WIPO Copyright Treaty, *supra* note 51, art. 5 (capitalization omitted).
318. *See id.* art. 14.11 (stipulating provisions regarding cross-border transfer of data and other information by electronic means).
producer's right, it would not be too far-fetched to suggest that the proposal for property rights in data will return in yet another form if the European Commission's proposal is rejected or fades away. However, if that proposal succeeds and slowly emerges on this side of the Atlantic, there is even more reason for policy makers and commentators to actively engage with those five sets of policy questions.

The scope and length of this Article do not allow for a further exploration of the four courses of action advanced in this Part. Nevertheless, it is the Author's hope that this outline of preliminary policy recommendations will provide useful insights into how we can critically examine and reconceptualize the European Commission's proposal for a new data producer's right for nonpersonal, anonymized machine-generated data. To some extent, the discussion in this Part highlights the immense complexities involved in developing a sound and holistic data governance regime. The discussion also demonstrates the interrelationship between different areas of law and policy.

VI. CONCLUSION

In October 2017, the European Commission advanced a proposal for a new data producer's right. This proposal invites us to think more deeply about the need to strengthen protection for nonpersonal, anonymized machine-generated data. To critically examine this proposal and to highlight the potential problems a similar proposal in the United States would create, this Article revisits the past developments surrounding sui generis database protection; examines our present technological, business, scientific, and personal needs; and identifies the complications the proposed right would generate in the law and policy arena.

In doing so, this Article underscores the immense challenges for developing a new data producer's right. Specifically, the proposal for developing this new right has generated an expansive list of policy questions and spill-over complications that policy makers will have to consider before creating that right. Policy makers should therefore undertake careful analysis and critical evaluation of this proposal. Where possible, they should also work preemptively to ensure the future development of a sound and holistic data governance regime.