IP Protection in the Data Economy: Getting the Balance Right on 13 Critical Issues

BY ROBERT D. ATKINSON | JANUARY 2019

Getting the debate over data and IP right is critical because regimes that tilt too far toward granting data rights run the risk of stifling needed data sharing, while regimes that tilt foo far in the other direction risk limiting incentives for data collection and innovation.

Intellectual property (IP) systems are designed in large part to provide adequate incentives for creators and inventors to invest in the production of novel ideas and content—while at the same time encouraging beneficial diffusion of knowledge. For example, by publishing patent disclosures, other inventors can learn about innovations; and by limiting patent terms to 20 years, other innovators can learn about and use patented innovations. As we move ever deeper into the data-driven economy, policymakers should take into account the need to maintain such a balance when considering the relationship between intellectual property rights and data.

The emergence of the data economy has led to a growing debate about data rights, related to both intellectual property and privacy. Getting the debate over data and IP right is critical because regimes that tilt too far toward granting data rights run the risk of stifling needed data sharing, while regimes that tilt foo far in the other direction risk limiting incentives for data collection and innovation.

THE GROWING DATA ECONOMY

We regularly hear about how data generation and use is exploding. In 2013, the Norwegian research organization SINTEF claimed that more data had been created in the previous two years than in the history of the human race—a figure that has been widely repeated.[1] However, much of this new data (IBM estimates we create 2.5 quintillion bytes of data daily) is a reflection of the massive increase in high-definition (HD) video. Even setting aside HD video, data continues to grow both in volume and importance to individuals and organizations. Recent technological improvements and advancements—

such as cheaper storage, faster computing, better algorithms, improved data sensors, and more robust communication networks—have made it easier and less expensive to collect, store, analyze, use, and disseminate data. These changes have led to the emergence of the data economy: an economy wherein success depends on how effectively organizations can leverage data to generate insights and unlock value. More and more organizations are using more and more data to help them provide goods and services.

Many companies now collect large amounts of data. GE and Siemens, for example, are actively working on services that would collect and analyze data from the machinery they sell.[2] IBM is integrating data from electronic health records, medical imaging, claims, and genetics to improve its Watson Health analytics service.[3] Automotive companies use connected vehicle data to improve their vehicles.[4] Supermarket chains have used data from loyalty-card schemes to offer personalized discounts; they now hope to use additional data from third parties to better time those promotions and compete with budget brands.[5] Whole industries, such as health care, agriculture, and consumer goods, are rapidly moving toward the collection of increasing amounts of data about their customers and products.[6]

Better use of data will be an important driver of economic and societal progress in the coming decades. The widespread adoption of data analytics and artificial intelligence is expected to contribute hundreds of billions of dollars to U.S. GDP in the coming years in sectors such as finance, transportation, and manufacturing, while unlocking new opportunities to improve outcomes in fields such as education and health care.[7]

HOW SHOULD WE THINK ABOUT DATA?

Using the right conceptual framework to understand data is important if we are to effectively frame IP and data issues—and conversely, ill-fitting conceptual frameworks can lead policymakers to bad conclusions. For example, many supporters of European government efforts to tax Internet firms (mostly based in the United States) on the basis of how much data they hold is justified in the framing that "data is the new oil."[8] If that were the case—if data really is as valuable, rivalrous, and excludable as oil—then the argument for taxing it would be strengthened, at least in the public's perception. But it isn't. In fact, most of the analogies that scholars and pundits use to explain why data is important end up having limited utility for policymaking purposes, and many are quite misleading.

Nonetheless, analogizing data to other things is a popular activity. Some have even equated data to bacon—because it has "sizzle."[9] The "new oil" framing is perhaps the most common fallback, however. As *The Economist* wrote, "The world's most valuable resource is no longer oil, but data."[10] But while it is likely true the value added to the global economy from data will be larger than that from oil, this cannot be the right analogy for the simple reason oil is a rivalrous and largely excludable good: If you have a barrel of oil, then I don't have it, and only one party can hold mineral rights to drill for oil in any particular area, whereas at least some data is by definition in the public domain (e.g., street addresses).

Others say, "data is oxygen."[11] But this cannot be the right analogy because while oxygen (at least the air we breathe), like data, is essentially non-rivalrous (two people can use it), it

is not excludable (no one can buy up a city's air and charge others to breathe it). But data can be excludable. For example, a company whose machines generate data can prevent others from using that data.

Another argument is that data is like infrastructure.[12] Reto Hilty, for example, wrote:

> Data might be grasped as (part of) the infrastructure of the digital economy: the "data-driven" economy. Following that logic has a particular significance because it leads us to the insight that "data" possibly are not just a private matter; at least public infrastructure obviously needs to be a concern of public authorities.[13]

But the problem with this analogy is that while some data (i.e., traffic data) can be integral to the public infrastructure, other kinds of data are private and only used by single organizations.

Ensuring the right conceptual framework for data is important if we are to effectively frame IP and data issues.

A somewhat better explanation is that data is like any other creative output, such as technical inventions or creative content. Both are non-rivalrous (multiple people can use a "recipe" for an invention) and can be excludable, as inventions and creative content are with patents or copyrights. However, this framing also does not quite fit for the simple reason that, while most data is useful, unlike an invention, much of it is obvious and not novel. Moreover, the reason inventions are patented and creative content is protected by copyright is, in order to be valuable, both have to be put out into the world through markets. Without legal protection, inventions and creative content would be more widely copied. But, the data brokerage business notwithstanding, the value of data does not usually result from its being sold; it comes from being used internally within a firm, where it is usually protected by technical means. As Duch-Brown, Martens, and Mueller-Langer wrote, "Unlike in copyright-protected media products that are meant to be distributed widely, data are not necessarily widely shared and may not need legal protection to make them excludable."[14]

But there is another reason why this framing does not quite work. The aggregation of data into what people term "big data" is often where the most value is created. For example, while having data about the location and speed of a particular car may be interesting, it is not all that valuable except to the person who is driving (and perhaps their family). But accumulating data from tens of thousands of cars in a metropolitan area and displaying it on a map (e.g., Waze) is incredibly valuable as it gives travelers and first responders real-time information about traffic conditions. In contrast, the value of a particular song does not increase when there are more songs; in fact, it likely goes down.

KEY POLICY ISSUES FOR DATA

So perhaps there is no really good analogy for data, and we just have to think about data as data. That said, what are some of most important issues related to intellectual property? This paper lists 13 key policy issues.

Issue 1: Policy needs to get the balance right between incentives for collection, curation, and analysis and benefits from widespread use of data.

In a review of the patent system, the Australian Law Reform Commission wrote:

> Patents promote innovation through the grant of limited monopolies, as a reward to inventors for the time, effort and ingenuity invested in creating new products and processes. The potential for financial returns adds an incentive to the traditional rewards of scientific innovation, such as academic recognition and promotion within research institutions. Without the incentive provided by patents, private investors may be reluctant to invest, resulting in greater calls on government funding or a failure to develop and exploit new technology.

But the commission also wrote:

> Patents promote knowledge sharing by requiring the details of the patented invention to be placed in the public domain in return for the exclusive right to exploit the invention. In the absence of this exchange, inventors might protect the details of new inventions through secrecy. The disclosure requirements of the patent system are based on the idea that "scientific and technical openness benefits the progress of society more than do confidentiality and secrecy."[15]

Much of the debate over patent policy is about where to draw the lines between protection and incentives for innovation versus less protection and greater sharing for learning. The same tension applies to how to think about ownership and usage rights for data.

Much of the debate over patent policy is about where to draw the lines between protection and incentives for innovation versus less protection and greater sharing for learning. The same tension applies to how to think about ownership and usage rights for data.

But as noted, data sharing, like patent sharing, is valuable because data is more useful when combined than it is in discrete form. This is why the European Commission wrote:

> Data is a non-rivalrous resource: it is possible for the same data to support the creation of several new products, services or methods of production. This allows any company to engage with the same data in different data-sharing arrangements with other big companies, SMEs and startups, or even the public sector. This way, the value resulting from the data can be exploited to the maximum.[16]

In addition to the information generated from large datasets, another reason data is more valuable when it is combined is machine learning algorithms are generally more valuable as the size of the training datasets becomes larger. As such, the fact that combined data is more valuable than the sum of individual but separate pieces of data suggests that any IP system for data should probably tilt toward data sharing. But that does not mean a data IP system should default to no IP rights, or even forced sharing. The costs involved with the collection, cleaning, and curation of data are often non-trivial, and when organizations that engage in such efforts lack exclusive rights to use that data, their incentives for collection, cleaning, and curation are diminished.

Issue 2: Data is non-rivalrous and often readily available.

The fact that data is non-rivalrous means that one company possessing data, even large amounts of it, does not mean other companies are precluded from also using the same or similar data. If a company knows a person's name, age, gender, income, and interests and

uses that information to effectively market to them does not mean another company cannot collect the same information about that person, either by purchasing it from companies that already have it or by finding ways to have them provide it. As Duch-Brown, Martens, and Mueller-Langer wrote, "The short history of the digital economy has so far shown that substitutes exist. Competitive advantage is not acquired by accumulating lots of data but rather by developing the organizational capabilities to make better use of data."[17]

This is not to say data cannot offer competitive significance for a given company. This is particularly true when assessing the temporal dimension of data. Some data is extremely valuable when it is generated, but rapidly loses value over time. For example, customers pay hefty fees to access business information from Bloomberg's proprietary terminals because getting this data before competitors can lead to advantage in the stock market. But the value of this same data 24 hours later is vastly reduced. Similarly, geolocation data is often more valuable in real time, for example, so that an application can serve up an ad relevant to where a person is or provide real-time traffic information. A day later that data often has much less value.

Yet it is true that much data is ubiquitous, low cost, and widely available. Government agencies offer large amounts of it for free.[18] An entire industry of data brokers makes a living collecting as much data as possible and selling it to companies that find it valuable. Other data, such as satellite and genomic data, might be expensive to acquire but not exclusive and still relatively cheap to share.

Although retaining data exclusively may confer a competitive advantage, it also imposes an opportunity cost on organizations in the form of lost revenue and fewer opportunities for innovation. This is why some companies are discovering the short-term benefits of hoarding data do not outweigh the long-term benefits of sharing it. For example, the Project Data Sphere helps pharmaceutical companies share cancer research data in the hope of accelerating discoveries.[19] It is likely the amount of lost revenue will increase with the exclusivity and strategic value of the data.

Issue 3: Who has rights to data?

To what extent, and through what means, should organizations have intellectual property rights to raw data, as opposed to databases? There are three main ways companies protect intellectual property: patents, copyrights, and trade secrets. Patents are not really an effective tool because data—at least most raw data—is almost always obvious and not novel. For example, as explained by U.K. attorney Jo Joyce, "[in Europe] raw machine-generated data are not protected by existing intellectual property rights since they are deemed not to be the result of an intellectual effort and/or have any degree of originality."[20] Moreover, like genes, at least some data is "a product of nature." The U.S. Supreme Court ruled that genes themselves cannot be patented because DNA is a product of nature.[21] The Court did rule that DNA manipulated in a lab could be eligible to be patented because the DNA sequences are created by humans and not found in nature.

So, what about data that is created by humans and not found in nature? A machine that generates data about its own workings is arguably creating data that is not found in nature. But as previously noted, the data might be novel, but it may be obvious in the sense that it takes little effort to create it. Moreover, as Hilty argued, "The establishment of legal exclusivity might produce unwanted, dysfunctional effects; instead of fostering the digital economy, certain business models might even be impeded."[22] In short, it makes little sense to establish a patent right to most data.

Copyright also is not the best means to provide IP protection for most data—or at least for data that required little creativity to generate. Certainly, digital content that involves creativity—an e-book, a digital file of a photograph, an MP3 music file—are and should be subject to copyright protection. But this is not because bits per se should be able to be copyrighted, but rather because it is the combination of bits that represents creative content.

Organizations should be permitted to use any means they wish to keep their data secure.

Trade secrets come closest to serving as the right means by which to protect many kinds of data, especially information organizations generate on their own or put considerable work into curating. It is not a violation for a company to possess the same data as another company as long as it collected or created the information on its own and did not take it from the other company. Some have argued that trade secrets are a bit of a kluge because some data is not necessarily kept secret—even though organizations do not want others to use it.[23] The point is not that the data is secret, but rather that the "owner" of the data wants to limit its use. Nonetheless, this comes closest to how to best protect data. Moreover, this does not preclude the use of contracts to protect the rights of a firm regarding its data if it shares that data with others, just as it would not preclude this in the case of copyrights. As Peter Bittner wrote, "Of course, contractual agreements between specified parties can be drafted to rule on the allowed use of data by the licensee."[24]

While trade-secret protection is a valid way to protect an organization's investment in data, there is also a supplemental protection related to computer hacking laws. Even more than traditional trade secrets, data is at risk from cyberattacks. Even if a company keeps data secret, that secrecy is gone if outsiders can break into its computer system and steal it. As such, stronger penalties against hacking and enforcement of computer anti-hacking and trespassing laws will help ensure stronger incentives for data collection.

Related to this is the ability of companies to use technical means to protect data. Encryption is the clearest example of this. When an organization effectively encrypts its data, both at rest and while in transit, its ability to keep its "ownership" and the value that comes with it increases significantly. However, some foreign governments have considered erecting regulatory limits to organizations' ability to use technical protection measures.

For example, Australia recently passed a law that requires companies to provide law enforcement and intelligence agencies with access to encrypted data.[25] Likewise, some in U.S. law enforcement have called for laws limiting the ability of organizations to securely encrypt data.[26] Even though law enforcement in democratic, rule-of-law nations has

legitimate reasons to access data, banning encryption is a not appropriate, in part because it raises systemic cybersecurity thefts, and serious law breakers will use encryption anyway.[27] Organizations should be permitted to use any means they wish to keep their data secure.

This issue is related to, but intellectually separate from, how and under what circumstances government should be able to access data. This is not an issue of balancing incentives for data collection and management with incentives for innovation from the wider use of data. It is an issue of balancing individual civil liberties with public interest in national security or crime prevention. In an increasingly digital economy, courts and legislatures will have to update rules and laws to address new applications of technology, such as storage of personal data in the cloud and the rise of connected cars with the ability to track location.

Finally, a related question is whether companies should have the right to sell data they collect. Organizations should be able to write clearly disclosed contracts that give them the right to sell data, provided they do not violate privacy laws. But this right to sell should not come with an exclusive right over the data that precludes others from using the same data without permission. If a company wants to sell personal information about someone (e.g., their birthday, etc.) it should have the legal right to. But it should not have the right to prevent other organizations from using that data if they can get it elsewhere. In this sense, ownership, particularly of personal data, is not the right concept or term. As Hilty wrote, "Already the term 'ownership' raises numerous questions. This becomes visible with the attempt to find alternative terms; they may exist—property, possession, exclusivity, control, sovereignty, responsibility—but they all (legally) mean something different, in particular in relation to data."[28] Rather, in most cases, a more appropriate framing is about rights of use.

Also, none of this suggests new laws could not be established, particularly at the sectoral level, defining access rights. For example, with the Health Insurance Portability and Accountability Act of 1996 (HIPAA), the U.S. health data privacy law, patients have a right to their data. Such rights could be established, either across the board or in particular areas, based perhaps on the sensitivity of the data. This gets to the next issue of whether individuals should have ownership or other rights to personal data about them.

Wolfgang Kerber argues that governments should not institute ownership rights for data because technical protection measures and contracts can provide adequate incentives for collection and management of data.[29] In contrast, Andreas Wiebe suggests that a data ownership right could be a coding right (first storage or recording of data) that requires novelty and registration, and should protect against copying but not against independent creation, for a maximum duration of five years.[30] But this would surely yield a vast array of legal squabbles as firms fought with each other over who first recorded someone's birthday or interest in tennis. It would potentially mean less—not more—data collected, if the first coder could prevent others from collecting the same information.

Issue 4: Should individuals own "their" data?

Data can either be related to a person or a "machine," although the dividing lines are not always strict. An example of machine data would be data from sensors in a farmer's field

reporting data on soil moisture, temperature, and growth rates. Personal data would such information as a person's date of birth, their emails, and their health records.

It is worth distinguishing between the four main kinds of personally identifiable information (PII). PII can be used to distinguish or identify an individual or can be linked or is reasonably linkable to that individual.[31]

The first category is observable information, which is personal information that can be perceived firsthand by other individuals. This category includes both observable personal information created by the individual about themselves, as well as observable personal information captured by a third party. An example of the former is personal correspondence, such as letters or emails a person has written. Examples of the latter primarily come from recorded media, such as video surveillance (e.g., CCTV camera footage), photographs (e.g., personal photos), or audio recordings (e.g., recording of a conversation). Media captures personal data in a way that, while recorded by a third party, any individual can observe it for themselves by looking at the photo, watching the video, or listening to the recording.

The second category of information is observed information, which is information collected about an individual based on a third party's observation or provided by the individual but does not allow someone else to replicate the observation. This data can encompass a wide variety of information that describes an individual, such as their basic information (e.g., place of birth, date of birth, etc.), physical traits (e.g., weight, eye color, etc.), personal preferences (e.g., likes and dislikes, political views, search history, reading habits, media consumption, etc.), social traits (e.g., degrees, religious affiliations, nationality, criminal history, etc.), family information (e.g., marital status, child information, etc.), employment information (e.g. job history, salary, etc.), biological conditions (e.g., sexual orientation, medical conditions, medical lab results, disability information, etc.), and geolocation information.

The third type is computed information, which is information inferred or derived from observable or observed information.[32] Computed information is produced when observable or observed information is manipulated through computation to produce new information that describes an individual in some way. For example, companies construct online advertising profiles for consumers based on many different sources of observed information, such as direct-mail responses, search history, and demographic information. Biometrics are derived through a computational process from scans of unique physical characteristics on a person's body. For example, during security screenings at airports, the Transportation Security Agency (TSA) uses backscatter X-ray machines to generate generic outlines of individuals' bodies, highlighting areas containing potential contraband in the images. Information in this category is primarily used to create value for the organizations that compute the information.

Finally, associated information is information a third party associates with an individual. Associated information, by itself and unlike the other three categories, does not provide any

descriptive information about an individual (i.e., it does not describe qualities about an individual). For example, a library card number alone does not provide any information about its owner. (Someone may be able to infer information about an individual based on the fact they have a library card, but the numbers on the library card itself generally convey no meaning about the individual.) There are many different types of associated information, such as government identification information (e.g., Social Security numbers, driver's license numbers, security clearances, etc.), contact information (e.g., name, home addresses, phone numbers, email addresses, etc.), device identifiers (e.g., IP addresses, MAC addresses, browser cookies, etc.), property information (e.g., land titles, vehicle registration numbers, etc.), online authentication information (e.g., screen names, passwords, security tokens, etc.), and financial information (e.g., bank account numbers, credit card numbers, insurance details, etc.).

Table 1: Examples of the four main types of personally identifiable information (PII)

Type of Information	Examples
Observable Information	Photographs, Videos, Emails, Recordings
Observed Information	Geolocation, Date of Birth, Search History
Computed Information	Advertising Profiles, Biometrics, Credit Scores
Associated Information	Social Security Numbers, IP Addresses, Land Titles

With observable information there often can be a clear owner, particularly when it involves an original work of authorship. Clearly, data that collectively forms a video, a sound recording an image, or text can be copyrighted.

With observed information, the issue of rights relates to the question of who contributed the "labor" to the data (and why). For data that is just "out there," such as someone's date of birth, there should no rights to control that data. The fact that a person was born in Canada is data about them, but the fact that this information can be easily obtained suggests they do not possess a right of control.[33]

But a lot of observed data is not widely available and is provided to other parties, such as social networks (e.g. posts/status updates). This involves user input, so clearly the user should have some rights. But when someone goes online, organizations (e.g., Facebook, Wikipedia, Amazon, etc.) possess information about them. Here companies should have rights to use the data, governed by their and government privacy policies, by the ability of users to opt out of secondary use, and by any privacy rules governing the rights of consumers to data deletion.

While it is also observed data, network traffic or SSID (Wi-Fi identifiers) broadcasts are fairly incidental "data exhausts." So, if a company is collecting that information, they

should have the rights to it. Then there are the middle ground cases, such as doctors' notes, wherein consumers should have either shared rights, or consumers should have primary rights, because they are paying the doctor for the service, which includes producing data about the patient.

Overall, to say that the companies a person shares information with have no legal rights to that data cannot be the right answer. This is especially true if the companies using that data curate it (fix errors, such as an incorrect birthday—born 1865, for example) or combines it in a way to make the overall database more valuable. But these companies should no more be able to own the rights to data they have about people, than those people should be able to own data about themselves. This is because, as U.K. attorney Jo Joyce, wrote, "No one can truly own a piece of data, the only thing that can be possessed is an aggregation or collection of such data, provided there has been a relevant investment in carrying out that aggregation or collection."[34] To be clear, an aggregation of bits into the form of a photo is something that should, and does, have intellectual property rights.

To say that the companies a person shares information with have no legal rights to that data cannot be the right answer.

With computed information, the rights should accrue to the party that performed the lion's share of work (e.g., computing). For example, when a company spends time and effort to create an advertising profile for an individual, that company should have the rights to that data, assuming they complied with all relevant privacy laws and regulations.

Finally, with associated information, the organizations that produced the information should have primary rights, but when it represents an individual, that individual clearly has usage rights.

But notwithstanding these differences, there have been broad calls for individuals to have ownership rights to their data. Every day, hundreds of millions of people go online to search the web, watch videos, read content, and catch up with friends—all without paying a single cent. Some critics deride this free ecosystem, claiming that not only are unsuspecting consumers "paying" for these services with their data, but they are getting a rotten deal. Many argue that individuals should have legal rights to "their" data and should be paid for it. For example, a recent article in the *Financial Times* entitled "Digital privacy rights require data ownership" is emblematic of the dominant framing: Ownership implies sovereignty over property, but unlike property, data can be copied, which means sharing it is not a zero-sum game like property transfers.[35]

These sentiments have been echoed by EU Commissioner of Competition Margrethe Vestager:

> Very few people realize that, if you tick the box, your information can be exchanged with others. … Actually, you are paying a price, an extra price for the product that you are purchasing. You give away something that was valuable. [36]

To fix this, there are several proposed solutions. Some argue that government should enact tougher privacy laws that give consumers the right to not contribute data or to have data on them held by third parties to eventually be deleted, and that in return, companies would

be required to provide the same quality of services at the same terms as for consumers who shared data. This would create a classic free rider problem, wherein if everyone insisted on this, the overall value for consumers would be less.

Others argue that Internet users are contributing valuable services when they input data and they should be paid for it, or that companies that collect this data are adding value in the nation they collected it and thereby should be subject to corporate tax obligations in that nation. The latest example can be found in the article, "A Blueprint for a Better Digital Society," by Glen Weyl and Jared Lanier, an Internet pundit. Not content with Internet users being able to access many online services like Bing and Twitter for free, they want online users to be paid in cash for the data they provide. Similarly, Herbert Zech has argued:

> Some good reasons exist for creating a new exclusive right to use data … for big data analyses pertaining to the person economically maintaining the machine. The reason is found not so much in an incentive to generate data or in the creation of a market for data (like in classical IP) but in ensuring a fair allocation of the profits generated by analysing the data. Instead of relying on existing factual ownership and secrecy, a clear property rule can provide the framework for a functioning data economy.[37]

The reality is users simply would not be paid that much for their data.

But the reality is users simply would not be paid that much for their data. For example, one estimate suggests Facebook could pay users around $15 a year for their personal data.[38] But this estimate is vastly overstated because it assumes companies can pass along all revenue and ignores all their costs. A more realistic model suggests users would get just a pittance. Google and Facebook, for example, earned about $28 billion in combined profits in 2017 and have around 4.6 billion users globally. If the payments to users were equal to half their profits, then each user would get just of $3 per year. One reason for this is that the value of personal data is often limited unless it is combined into larger data sets where analytics can be performed. For example, the value of a person's genomic information is relatively low unless that data can be combined with health information and with a large number of other people's data.

Contrast this with the value users get from free online services. Economists Jay Corrigan and Matt Rousu found that on average American Facebook users would have to be paid more than $1,000 per year to give up Facebook.[39]

However, data is not like cash, and enacting laws and regulations based on this misconception would both harm the digital economy and make the lives of digital consumers considerably worse. As Duch-Brown, Martens, and Mueller-Langer wrote in a report for the European Commission, "Advertising is mostly seen as a pro-competitive effect and not harmful from a competition perspective. The absence of monetization would reduce the volume and increase the cost of online services and reduce competition in product markets."[40] Moreover, Weyl and Lanier's proposal would significantly increase transaction costs associated with sharing data, thus substantially reducing consumer value. Not only would this proposal likely require obtaining consent from all users, it would also

require contracts between users and intermediaries, along with financial information for payments, thus creating an epic quagmire of paperwork. While there is significant value in large datasets, the marginal costs of each additional data point may be minimal, and not outweigh the transaction costs. Thus, it may be impractical to share data with separate ownership rights assigned to consumers given the transaction costs involved—as opposed to data brokers who can license large datasets, thereby creating a significantly lower per-user transaction cost.

Higher transaction costs would also make it unfeasible for many companies to collect and use consumer data, which could lead to fewer online services available to consumers. To ensure "a better digital society," companies should continue to be allowed to decide the best Internet business models based on consumer demand. Data is neither cash nor a commodity, and pursuing policies based on a misconception of this will damage the digital economy and make the lives of digital consumers considerably worse.

There are good reasons to believe market forces will ensure fair data relationships.

This is because the exchange of data is a fundamentally different exchange of value than other transactions. Data is non-rivalrous: Many different companies can collect, share, and use the same data simultaneously. Similarly, when consumers "pay with data" to access a website, they still have the same amount of data after the transaction as before. As a result, users have an infinite resource available to them to access free online services. In other words, if someone gives you $10, they have $10 less. But if they tell you they are a basketball fan, then you both know that information. Sharing their data does not preclude them from sharing the same data to access any number of services.

There is another reason why it would be a mistake to give consumers exclusive rights to the data they share with other organizations. Ad-supported digital services turn data into value by functioning as two-sided markets that connect consumers and advertisers. Users get access to a free service, and advertisers get access to an audience for their ads. In most cases, the advertisers do not even know which users see their ads, only that the ads are placed in front of a targeted group of people, such as people who live in a particular location or have specific interests.

Moreover, the sharing of data, even with commercial companies, creates significant societal value. As Geoffrey Manne and R. Ben Sperry argued:

> The size of a database (i.e., the number of consumers on whom data is collected) doesn't seem like a particularly relevant aspect of product quality in and of itself, and for each consumer the "problem" of a large concentration of information being accumulated in a single company is seemingly insignificant. Meanwhile, to the extent that collection of data from more consumers is a function of increasing network effects, such accumulations of data are almost certainly more likely to correlate with improvements in product quality rather than degradations.[41]

This is not to say that companies should not be allowed to or even encouraged to come up with business models wherein they enable consumers to monetize their personal data. For example, Microsoft is purportedly looking to provide users with access to their own

personal data bank through its "Project Bali" effort. If the business models work for this or other related efforts, this practice will surely grow.[42] But that is quite different from government mandating that individuals have a monetary right to compensation.

Issue 5: Who should own private, non-PII data?

Increasing amounts of data are being produced or captured by machines. For example, global mining company Rio Tinto has created its "Mine of the Future" program to "identify the size, location and quality of ore" by aggregating the data it collects in real time. Rio Tinto collects this data from both the trucks and the drills it uses in its mines all around the world and processes it at its Processing Excellence Centre (PEC). Its manufacturing operations are slowly becoming "intelligent," with more and more machines being sensor-based and connected, enabling real-time analytics to be run on machines and even in whole establishments.[43] One issue related to such machine data is who owns the data? For a factory, is it the manufacturer who owns the machine? Is it the machine maker who sells the machines to the manufacturer? Is it the third-party system integrator who connects a company's machines?

Higher transaction costs would also make it unfeasible for many companies to collect and use consumer data, which could lead to fewer online services available to consumers.

A number of nations are focusing on this issue. For example, the European Commission is worried that big companies selling machines or software will have market power and force customers, particularly smaller companies, into relationships wherein the machine seller owns the data, not the machine buyer. Similarly, the Japanese government has also been considering these questions. Underlying both of their concerns is the worry that U.S. companies, particularly "big-tech" companies, will use their market power to extract unfair data concessions. As such, the European Commission has established a working group of experts to examine this issue. The Commission aims "to ensure fair and competitive markets for Internet of Things (IoT) objects and for products and services that rely on non-personal machine-generated data created by such objects."[44] The Commission also suggests a number of draft principles for companies to consider when drafting relevant contracts. These concerns are also expressed in the United States, often by organizations worried about losing control of data.

But these concerns appear to be misplaced, or at least premature. First, the norm, at least presently, is for machine buyers to own the data, at least for commercial and industrial applications. Most companies do not want the data going outside of their control. To be sure, some machine builders and system integrators have created lightweight, secure, remote VPN access that allows a company to access data at a manufacturing site if the company requests it; just like a PC user can let a remote help desk access their PC to fix any problems. But like a help desk assisting someone with their PC, when a remote factory task is complete, the connection is closed. There is a second reason why the business model is likely to be one in which machine owners own the data. Pushing all of the data out to a third party increases the threat from hacking, compared to just one company holding it. It is also sometimes not practical because of the large quantity of data involved. Some machines have I/O backplanes that operate at 1 gigabyte (GB) per second and there can be thousands of these in a given factory.

Moreover, there are good reasons to believe market forces will ensure fair data relationships. In few industries do machine sellers have a monopoly. Therefore, they have an incentive to provide the kinds of products and services customers want. Moreover, bad publicity from "unfair" data practices can be real and are something most companies seek to avoid. We see this dynamic in the agricultural sector, where concerns by U.S. farmers that big agricultural firms, such as John Deere, would control data from precision and smart agricultural systems. In 2014 however, working with farm organizations, John Deere and a number of other large agricultural companies signed on to a set of data principles, including one on ownership. The principle states:

> We believe farmers own information generated on their farming operations. However, it is the responsibility of the farmer to agree upon data use and sharing with the other stakeholders with an economic interest, such as the tenant, landowner, cooperative, owner of the precision agriculture system hardware, and/or ATP [agricultural technology provider] etc. The farmer contracting with the ATP is responsible for ensuring that only the data they own or have permission to use is included in the account with the ATP.[45]

But even with these technology and market forces keeping the data in the hands of the machine owners, there may be cases where machine sellers have a business model that involves the collection of data from multiple factories and companies, with their permission, and aggregate it and perform machine analytics on it to provide back to each manufacturer information that enables them to improve performance, not only of the machines but of the overall agricultural system. As such, government regulations prescribing a particular business model or specific contract language run the risk of limiting, not advancing, innovation, especially if they limit data aggregation by machine or system sellers.

Issue 6: Forced sharing.

Even if the IP measures for data protection cannot be as robust as for patents or copyrights, there can still be protection, with trade secrets, encryption, and other means. However, just as some governments impose compulsory licenses on patents or decriminalize Internet piracy by users, some pundits have advocated, and some governments have considered, forced sharing of business data. However, policymakers need to distinguish between forced sharing of proprietary information only the company has and forced sharing of more public information (e.g., customers' smart meter data).

Viktor Myer-Schonberger and Thomas Ramge have called for a regime of forced data sharing, in part to respond to supposed competition concerns.[46] Arguing that big data gives some firms an unfair competitive advantage, they portray their proposals as less onerous than an antitrust breakup of big technology firms. Rather, they propose a data-sharing mandate wherein government requires big data firms "to share anonymized slices of data they collect with other companies."[47] However, the claims that big data, especially collected PII, gives companies an unfair competitive advantage are weak.

Regardless, mandatory data sharing would be a complex and onerous process to regulate, in part because government would determine what share of data must be disclosed based on a firm's market share. Moreover, firms would have to provide data at random "to prevent companies from gaming the system." But who would decide whether the data shared is in fact random? Moreover, as Mark MacCarthy wrote, there would be significant privacy implications if the data shared were about persons.[48] While Myer-Schonberger and Ramge proposed the data be anonymized, there is a risk of reidentification if the receiving company also has data on the same individuals. Moreover, just as forced licensing of drugs reduces the revenue and incentives for innovators, forced data sharing would do the same. It costs money to collect, clean, organize, and maintain data. Forced sharing assumes data and its collection, storage, and management are free.

When it comes to competition policy, the focus should be on anticompetitive conduct and not on structural issues, such as how much data a company holds.

To be sure, in some cases, regulators have required companies to make data available to competitors.[49] This typically occurs in the context of a proposed merger the regulators believe would result in the new entity having a dominant position in a particular data market. In order to encourage new entrants or protect the position of existing competitors, competition agencies can require the merged entity to sell data to rivals at a market price. So far regulators have largely resisted calls to use this power to introduce greater competition in the absence of a merger or specific anticompetitive behavior.

Continued restraint is wise because mandatory sharing might actually increase privacy and data-security concerns in cases involving personally identifiable information. Because the incumbent would have limited power to attach appropriate security and use requirements on its rivals, firms that lack either the capacity or incentive to impose high data standards might end up possessing the data. Regulators would have to spend significant resources in order to implement and enforce any restrictions from outside. And privacy restrictions on the market for data brokers would make it harder for new companies to gain a foothold in downstream markets.

Issue 7: Competition policy and data ownership.

Forced sharing is often raised in the context of competition policy.[50] A number of commentators have begun to argue that in the case of companies aggregating large amounts of data, competition policy should be extended to incorporate concerns about the collection and use of data beyond clear examples of anticompetitive behavior.[51] The general argument is that the mere act of collecting large amounts of data, such as the vast quantities of personal data collected by social-networking platforms, search engines, and e-commerce sites, gives companies an unfair competitive advantage, and that competition policy needs to incorporate this analysis, particularly to help smaller firms.

For example, in a recent speech on data and competition, the European Commission's Commissioner for Competition Margrethe Vestager stated:

> It's possible that in other cases, data could be an important factor in how a merger affects competition. A company might even buy up a rival just to get hold of its data, even though it hasn't yet managed to turn that data into money. We are therefore exploring whether we need to start looking at mergers with valuable data involved, even though the company that owns it doesn't have a large turnover.[52]

Similarly, a recent European Commission report worried:

> Where business models of entire ecosystems of SMEs [small and medium enterprises] are dependent on access to a small number of online platforms, or where platforms have access to datasets of unprecedented size, new asymmetries may be created. In such situations, some suppliers to platforms can be disproportionately exposed to potentially unfair trading practices, even in the absence of established dominance of a platform.[53]

To date, U.S. and European regulators have not adopted this line of reasoning—nor should they. While it is true that data can be used in anticompetitive ways, competition policy is capable of dealing with such abuses. In fact, when analyzing allegations of such behavior, it is often helpful to imagine whether agencies would object if the activity complained about involved some input of critical importance other than data. This helps clarify whether the threat to competition is truly due to control of an important resource or to ungrounded fears about the uniqueness of data.

Advocates for intensifying competition policy cite a variety of flaws and potential abuses in the current system. However, defenders of the current approach seldom argue that there can be no anticompetitive behaviors when it comes to data. Rather, they admit that, in some cases, data use could trigger competitive concerns. What defenders do argue is that, when it comes to competition policy, the focus should be on anticompetitive conduct and not on structural issues, such as how much data a company holds.

Collecting large amounts of data does not by itself represent a threat to competition. Although use of data might in specific circumstances justify regulatory intervention, in most cases the acquisition and use of data do not reduce competition, and the existing legal framework, including traditional interpretations of existing statutes, gives competition and data protection regulators all the flexibility they need to protect markets and consumers. On the contrary, large amounts of data, including personal information, are increasingly a vital input for some of the economy's most important innovations, including online platforms, medical diagnoses, digital assistants, language translation, urban planning, and public safety.

Proponents of maintaining the current approach to competition policy point out that these effects also deliver tremendous value to consumers and society, so regulators should be careful in regulating them.[54]

With respect to data-intensive companies, there are many arguments for bigness, but this bigness benefits society. In many industries, marginal costs increase with higher production, in which case the supply curve slopes up. In contrast, for most information-

based industries, production costs fall dramatically to a point where the marginal cost is almost zero. As a result, these companies are able to lower the prices they charge users, at least until they attain a certain level of volume. For similar reasons, some people worry that as companies gain more access to information, they will be able to establish a dominant position because they will have achieved significant economies of scale.[55]

On the demand side, network effects ensure the value to each user rises as more users use the same service. The first Harvard students to use Facebook benefitted from it. But this benefit increased dramatically as the first billion users joined. Again, these effects probably trail off after a certain point, but the value of Facebook would be diminished if half the people were still on Myspace and not Facebook. Economies of scale and network effects both increase consumer welfare by lowering costs and increasing value. And they do not necessarily ensure lasting market power.

A similar concept holds for users. If users had to pay for each site, then the payments to one site would preclude spending the same money on another site. But users can furnish basic information, such as their email address, age, and shopping habits to as many sites as they want without diminishing their income. Users do face a time constraint in that it is difficult to spend the same hour on both Facebook and YouTube. But even this constraint is minimal because if a better product or service appears, consumers can shift their use of future hours very readily.

In a recent article, economists Anja Lambrecht and Catherine E. Tucker examined big data using a resource-based view of the firm, which holds that for a resource such as data to provide a company with a competitive advantage, it must be inimitable, rare, valuable, and non-substitutable.[56] They concluded:

> The unstable history of digital business offers little evidence that the mere possession of big data is a sufficient protection for an incumbent against a superior product offering. To build a sustainable competitive advantage, the focus of a digital strategy should therefore be on how to use digital technologies to provide value to customers in ways that were previously impossible.[57]

Likewise, Duch-Brown, Martens, and Mueller-Lanager have argued that "the short history of the digital economy has so far shown that substitutes exist. Competitive advantage is not acquired by accumulating lots of data but rather by developing the organisational capabilities to make better use of data."[58]

Even if the possession of large amounts of data were necessary for an entrant to compete successfully, that would not necessarily constitute an unfair competitive advantage. Many industries have high start-up costs. We do not say that Ford and Daimler have an unfair advantage just because companies must first build an expensive factory before they sell a single car. Nor does amassing a large number of workers represent a barrier to competition, even though these same workers are not available to competitors. Some things are just inherent to the business of offering customers a valuable product. In contrast, collecting data can be relatively cheap, and the data remains available to others.

Although barriers to entry are an element of antitrust analysis, these barriers can be less imposing than they look. Companies have often been able to overcome high upfront costs, provided they have a compelling business plan for eventually earning enough profits to deliver an appropriate risk-adjusted rate of return. An entire ecosystem of angel investors, incubators, and hedge funds exists to invest in promising young companies capable of growing rapidly. Although funding is often a challenge, the larger bottleneck remains a lack of innovative and workable ideas.

Finally, it is important to recognize that different companies have different strategies and business models around data. For some companies, their competitive advantage is the algorithm; for others, including some that are making their algorithms open source, it is the data. For the former case, IBM is training its cognitive computing system, Watson, to help analyze medical information, including the discovery of new drugs for immuno-oncology.[59] To do this, it needs lots of data. But the data would be much less valuable without Watson's sophisticated artificial intelligence capabilities. Sometimes these algorithms are protected as intellectual property, but that does not prohibit competitors from trying to write better ones. And sometimes these algorithms are made public.[60] For example, Google published the source code for its artificial intelligence engine, TensorFlow, to encourage others to find uses for and ways of improving it that Google might not have previously considered.[61] But even the best algorithms can be defeated by poor business strategy. As an example, one ex-executive attributes the fall of Myspace largely to poor business decisions.[62]

There can be cases wherein governments should, largely through competition policy, require open data access through open application programming interfaces (APIs).

Issue 8: Forced access.

Over the past few years, some scholars, advocates, and policymakers have argued that businesses that possess large quantities of data, such as social media companies, present inherent competitive concerns. As discussed, these concerns are misplaced for a number of reasons. But in some industries and markets, a small number of firms have exclusive access to particular datasets, and they exploit their market power to limit access to that data through both technical and administrative means without any legitimate business justification. This type of anticompetitive behavior limits innovation and hurts consumers, and when these problematic practices occur, policymakers should intervene.

In particular, there can be cases wherein governments should, largely through competition policy, require open data access through open application programming interfaces (APIs). For example, businesses and their associated industry associations, in the real estate, financial services, and air travel industries, have taken steps to limit third-party access to their data in ways that restrict competition, reduce market transparency, and harm consumers. In the banking industry, for example, some traditional financial institutions, such as banks and brokerage firms, prevent financial data aggregators, such as Yodlee and Plaid, from accessing customer account information via financial institutions' online services or APIs. Some financial institutions have an incentive to block financial data aggregators from downloading their users' data because these services are used by many fintech companies—businesses using innovative technology to improve financial services—

to show consumers ways to reduce the fees they pay for financial services. Without the data, these fintech businesses have a much harder time providing online tools to allow users to more effectively manage their finances. But the principle here is consumers and anyone they designate should be able to access their information.

This brings up a related issue of how to access that information. Using a banking example, if a consumer sends a third-party financial app an image of their bank statement for the company to manually input into its database, there is nothing in current law—nor should there be—to prevent that transaction from happening. But what is the practical difference between that and the consumer giving the third-party financial app their login credentials to the online bank account and having an app employee read the statement and type the information into its system? In reality, there is no difference. Now what about this example? The customer shares the login information and instead of the app employee logging in, the app's computer logs in automatically and copies the customer's information (and only that information) into its database. Again, I would argue there is nothing conceptually different about this.

This process uses what are called open APIs. APIs are software functions that allow developers to access data stored in computer systems in a prespecified, machine-readable format. APIs are routinely used within organizations, but open APIs allow third-party access to information as well. Providing third parties with access to this information serves consumers by increasing market transparency and allowing them to make more informed choices.

So, in what situations are mandated open APIs warranted, and where would they harm innovation and competition? Certainly, one area where they are warranted is where they enable the consumer or their agent to access their data. This is the case in banking, real estate, and electric utilities (e.g., data on customers' electric usage obtained from a smart meter). The advantage of open APIs in these situations is it enables more innovation and competition by letting other companies add value to people's data.

A second area where they may be warranted is where the data is already available to consumers and competitors, but is in a different form. For example, in the air travel industry, some airlines, such as Delta and Southwest, block certain third-party sites from posting flight availability and pricing information on their sites. Airlines have also targeted both specific online travel agencies (OTAs) such as BookIt.com and OneTravel, and metasearch engines such as TripAdvisor and Hipmunk, which let consumers easily compare fares across multiple airlines. But this information is not a trade secret, and it is not encrypted or otherwise protected. In fact, consumers and competitors could go online and manually find all this information, however time consuming it would be. Requiring open APIs would spur more transparency and competition.

Issue 9: Text and data mining.

Another related issue is that of automated text and data mining systems that more efficiently collect and parse data. We see this issue played out in the European Union with the European Commission's Proposal for a Directive on Copyright in the Digital Single Market, which aims to boost research and innovation by allowing research institutions to carry out text and data mining on lawfully accessed copyright-protected works. However, this exemption should apply to everyone, as text and data mining is either copyright theft or it is not, regardless of who is doing it. Moreover, it was not obvious whether text and data mining of intellectual property necessarily constituted a breach of copyright law in the first place. The exemption therefore reinforces the prohibition, which until now was not explicit.

European researchers have become frustrated in recent years by the restrictions EU copyright laws put on their freedom to use text and data mining—two automated techniques for analyzing data—on resources they can legally access and analyze with nonautomated means. As part of its recent proposals to reform copyright laws, the European Commission has recommended lifting these restrictions, but only for academics. For scholars and scientists, access to the rigorously scrutinized work of their peers, such as academic journals and databases, has always been a vital resource. Researchers who subscribe to these sources can explore them using traditional keyword searches and meta-tags predefined by publishers, but that has serious limitations. Manually reviewing all of these sources is a slow and tedious process, the results of which are often inaccurate and incomplete. Text and data mining are a powerful tool that allows researchers to plow into texts and datasets and interpret minute details.

However, the use of data mining on copyrighted material often falls afoul of existing intellectual property laws because the technical process involves extracting data from its original source and copying it into another database for analysis. In this case, Europe's proposed exemption is reasonable because it creates a special dispensation for data mining and does not alter other laws that prohibit the unauthorized extraction or reproduction of copyrighted works. After all, there is nothing illegal about mining databases manually; this technology only automates the process. A researcher could legally sift through many thousands of published works, note their findings with pen and paper, and then analyze the assembled notes. But since this process is legal with a pen and paper, it should be legal using new digital technologies—including data mining technologies—and it should be legal for anyone, not just academics. Copyright law should allow publishers to set the subscription fees for access to their content, prohibit unauthorized reproductions of their content, and receive appropriate compensation. But it should not require people with lawful access to content, such as paid subscribers, to seek approval from publishers for using automated research methods.

Issue 10: Database protection.

Automated text and data mining touches on the issue of database protection. In the United States, databases can be protected by copyright as long as there is some creativity and work involved in their construction. In contrast, in the EU databases have sui generis protection, even databases that are simply collections of facts. The protection, though, applies only when "there has been a qualitatively and/or quantitatively substantial investment in either the obtaining, verification or presentation of the contents."[63]

In the United States the realtors' Multiple Listing Service database of homes for sale is protected by copyright, even though some argue it should not on the assertion that it is simply a collection of facts. But even here the possession of copyrights does not mean another party cannot collect the same information and create a competing database. Moreover, the data itself is not copyrightable, although data complications can be, particularly when there has been work involved in improving the data or its presentation.

Issue 11: Government access to and use of data.

Much data, both personal data and machine data, becomes much more valuable when it is aggregated, and often that aggregated data can have important uses for society as a whole.

Another issue relates to government use of private data. Much data, both personal data and machine data, becomes much more valuable when it is aggregated, and often that aggregated data can have important uses for society as a whole—and governments in particular. Data to track disease breakouts, energy usage, air quality, educational results, and a host of other areas can play key roles in advancing government missions.

So how does government get this data? In some cases, government is able to obtain it through traditional means such as census surveys or collecting weather data through national weather service agencies. In other cases, they can create apps and encourage citizens to contribute data. For example, the Boston city government created an app called Street Bump that detects potholes and other bumps in the road from the sensors in people's smartphones as they travel on the streets of Boston. Users voluntarily agree to have their phone be an anonymous sensor.[64] To the extent government has this nonproprietary data, it should release such datasets through open APIs to let others use it. Data should be open by default, and not released only when required—and remain in the public domain.

In many cases, public welfare would be advanced if government could gain easier access to public data for key public tasks. Again, the challenge here is to enable data sharing to increase its value while minimizing limits to innovation. In contrast to compulsory licensing for medicines where it clearly reduces revenue and incentives for innovators, similar regimes for certain kinds of data can be imposed without limiting incentives for original data collectors to collect, curate, and add value to data, particularly if the government does not share this data with competitors.

The European Union is also considering new rules that would require some companies doing business with European governments to share data pertinent to public undertakings (such as transportation, energy management, health care, etc.). For example, a private company providing bus services under license from the local government or transport regulator would have to make available the data created as part of that undertaking.

Similarly, the French government has called for legislation to mandate repurposing both public- and private-sector data to enable public-interest uses of artificial intelligence by government and others, depending on the sensitivity of the data. For example, public health services could use data generated by Internet of Things (IoT) devices to help doctors better treat and diagnose patients. Repurposed data held by private companies could be made publicly available, shared with other companies, or processed securely by the public sector, depending on the extent to which sharing the data presents privacy risks or undermines competition. The French report suggests the government would not require companies to share data publicly when doing so would impact legitimate business interests, nor would it require any personal data be made public. If wider data sharing harms a company's commercial interests, it would be appropriate to give public authorities access to only the data the results of which they would publish, but not the raw data. And where the stakes are lower, companies could be required to share the data more widely, to maximize reuse. The risk of these European proposals is the definition of public purpose being too broad, and the definition of legitimate business interests being defined too narrowly, and companies being forced to share proprietary information in ways that limit business interests.

Issue 12: Conflicts between international regimes.

An additional factor for policymakers to consider is international differences in data regimes. Like inventions and copyrighted content, much data is cross-national in nature, including in such "traditional" industries as mining, retail, banking, and manufacturing.[65] Yet, nations are engaged in a host of policy areas that can limit data sharing across borders, including privacy laws in the EU General Data Protection Regulation, data localization laws, rules regarding government access to data, and restrictions on data contracts.[66] Likewise, different subnational governments have different rules regarding data.

As a representative of one multinational company wrote:

> Depending on their operational coverage, global companies like ABB need to deal with many different and divergent legislative regimes with diverse requirements and legal security. Given the need for data protection and the level of synchronization of national laws at present, international companies are obliged to adjust their business models to safeguard nationally available protection, and so their investments in the digital world.[67]

Because much data naturally flows across government borders, bringing more consistency and a shared legal and regulatory framework will be important. While it is unlikely and undesirable to harmonize national privacy laws, it is possible and desirable to enact trade rules limiting restrictions on cross-border data flows. This is easier than many policymakers might believe because data protection rules follow the data, regardless of where it is ultimately processed. At the same time, we should consider the analogue of the WTO TRIPS (Trade-Related Aspects of Intellectual Property) agreement for data, not so much to enshrine data rights, but to ensure data rules are harmonized across borders to the extent feasible.

Issue 13: The political economy of IP and data.

As with any IP issue, there is a political economy for data ownership. At the risk of overgeneralizing, the left sees data ownership differently depending on whether the owner is the consumer or the company. Generally, they believe consumers should be given an ownership right to their data, even if doing so would result in less data innovation (e.g., data analytics conducted on diseases). In contrast, they believe companies should face limits to data ownership and be subject to forced sharing of private company data even if it would limit incentives to collect, process, and analyze data. In contrast, free-market conservatives are generally skeptical of the idea consumers should own personal data, and argue companies should have rights to the data they possess.

A second political-economy component comes from other nations who see American "big tech"—or "GAFA" (Google, Apple, Facebook and Amazon), as some in Europe refer to them—as monopolists controlling vast amounts of data they will use to extract unfair deals with these nations' domestic companies. For example, the European Commission has talked about enshrining the right of data ownership as a way to keep American tech giants from owning "their data" and for justifying taxing American data companies. An EU report states, "In order to extract the maximum value from this type of data, market players need to have access to large and diverse datasets. However, this becomes more difficult to achieve if the generators of the data keep it to themselves, and the data is consequently analyzed in silos."[68] This seems implicitly directed against large American data companies. The report also concluded that "the lack of a legal environment adapted to the trade of data within the EU may contribute to insufficient access to large datasets, possible entry barriers to new market entrants, and stifling effects on innovation."[69] What is ironic, of course, about this is that at one level, Europe, through the GDPR, wants only consumers to own data, not European companies. But they can agree that at least U.S. tech companies should not control data.

Regulators should be concerned about stifling the large social value created by the gathering, analysis, and sharing of data. Innovation often depends on it.

CONCLUSION

Regulators should be concerned about stifling the large social value created by the gathering, analysis, and sharing of data. Innovation often depends on it. Moreover, if regulators began preventing companies from acquiring large amounts of data, it would delay or prevent many important technological advancements. For example, Tesla's self-driving technology (which faces increased competition from Google, rival carmakers, and others), IBM Watson's ability to diagnose medical illness, and the Weather Company's weather predictions would all be impossible without massive amounts of data. Data is also how Google often knows what you are searching for before you finish typing it in, how Facebook connects you with lost friends, and how Waze calculates the best route for you to take—all conveniences consumers already take for granted.

While data is incredibly valuable, so too is sharing and combining. It is very easy to reduce total welfare by overestimating the threats of data gathering and dismissing the public benefits of new products that do not yet exist. The fact is many data-rich companies offer free or low-cost services that are extremely valuable to billions of people, most of whom

have a pretty good idea of what data they are providing companies and how it might be used. Existing laws can deal with clear abuses. However, regulators can do a lot of damage by restricting the gathering and use of data in pursuit of preferences that are shared by only a minority of the population or by mandated forced access of data to supposedly level the playing field for companies that would prefer not to pay for data access.[70]

ENDNOTES

1. Ase Dragland, "Big Data—For Better or Worse," *SINTEF,* May 22, 2013, https://www.sintef.no/en/latest-news/big-data-for-better-or-worse/.
2. "Siemens and General Electric Gear Up for the Internet of Things" *The Economist*, December 3, 2016, http://www.economist.com/news/business/21711079-american-industrial-giant-sprinting-towards-its-goal-german-firm-taking-more.
3. Steve Lohr, "IBM Buys Truven for $2.6 Billion, Adding to Trove of Patient Data," *The New York Times*, February 18, 2016, https://www.nytimes.com/2016/02/19/technology/ibm-buys-truven-adding-to-growing-trove-of-patient-data-at-watson-health.html.
4. Jonathan M. Gitlin, "Audi Wants Its Connected Cars to Improve the Breed" *Ars Technica,* September 23, 2016, https://arstechnica.com/cars/2016/09/personal-assistants-and-data-analytics-the-future-of-audis-car-ux/.
5. Rachel Willcox, "Big Data, Empty Bellies: How Supermarkets Tweak Prices Just for the Sake of YOUR LOVE," *The Register,* February 5, 2015, http://www.theregister.co.uk/2015/02/05/big_data_tech_weapons_in_supermarket_price_wars/.
6. "Will Artificial Intelligence Help to Crack Biology?" The Economist, January 7, 2017, http://www.economist.com/news/science-and-technology/21713828-silicon-valley-has-squidgy-worlds-biology-and-disease-its-sights-will; Steve Lohr, "Medicaid's Data Gets an Internet-Era Makeover," The New York Times, January 9, 2017, https://www.nytimes.com/2017/01/09/technology/medicaids-data-gets-an-internet-era-makeover.html; Vonnie Estes, "How Big Data is Disrupting Agriculture from Biological Discovery to Farming Practices," Agfunder News, June 9, 2016, https://agfundernews.com/how-big-data-is-disrupting-agriculture-from-biological-discovery-to-farming-practices5973.html; "A New Industry Has Sprung Up Selling 'Indoor-Location' Services to Retailers," *The Economist*, December 24, 2016, http://www.economist.com/news/business/21712163-there-money-be-made-tracking-shoppers-paths-inside-stores-new-industry-has-sprung-up.
7. Susan Lund et al., "Game Changers: Five Opportunities for U.S. Growth and Renewal" (McKinsey Global Institute, July 2013), http://www.mckinsey.com/insights/americas/us_game_changers.
8. Joe Kennedy, "Resist Unilateral EU Efforts to Change International Tax Law for Corporations," *ITIF,* July 26, 2018, https://itif.org/publications/2018/07/26/resist-unilateral-eu-efforts-change-international-tax-law-corporations.
9. Joe Francica, "Data is the New Bacon," *IBM Business Analytics Blog,* October 18, 2018, https://www.ibm.com/blogs/business-analytics/data-is-the-new-bacon/.
10. "The World's Most Valuable Resource is No Longer Oil, But Data," *The Economist,* May 6, 2017, https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data.
11. Adam Bowen, "Data Is the Oxygen. It's What Moves the World," *DELPHIX,* August 24, 2017, https://www.delphix.com/blog/data-oxygen-moves-world.
12. National Infrastructure Commission, "Data As Infrastructure" (2017), https://www.nic.org.uk/wp-content/uploads/Data-As-Infrastructure.pdf.
13. Reto Hilty, "Big Data: Ownership and Use in the Digital Age," Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data 5, 87, https://www.ictsd.org/sites/default/files/research/ceipi-ictsd_issue_5_final_0.pdf.
14. Duch-Brown, Martens, Mueller-Langer, "The Economics of Ownership, Access and Trade In Digital Data."
15. Australian Law Reform Commission, The Patent System, https://www.alrc.gov.au/publications/2-patent-system/economic-benefits-patent-system.

16. Data Policy and Innovation (Unit G. 1), Guidance on Private Sector Data Sharing, (European Commission, August 2018), https://ec.europa.eu/digital-single-market/en/guidance-private-sector-data-sharing.
17. Duch-Brown, Martens, Mueller-Langer, "The Economics of Ownership, Access and Trade In Digital Data."
18. Daniel Castro and Travis Korte, "Open Data in the G8: A Review of Progress on the G8 Open Data Charter" (Information Technology and Innovation Foundation, Center for Open Data, March 17, 2015), http://www2.datainnovation.org/2015-open-data-g8.pdf.
19. A.K. Green et al., "The Project Data Sphere Initiative: Accelerating Cancer Research by Sharing Data," *The Oncologist 20*, no. 5 (May 2015): 464-e20.
20. Ibid, p. 10.
21. "Can Genes be Patented?" U.S. National Library of Medicine, September 25, 2018, https://ghr.nlm.nih.gov/primer/testing/genepatents.
22. Reto Hilty, *Big Data: Ownership and Use in the Digital Age.*
23. Claudia Jamin, "Managing Big Data in the Digital Age: An Industry Perspective," Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data 5, 150, https://www.ictsd.org/sites/default/files/research/ceipi-ictsd_issue_5_final_0.pdf.
24. Peter Bittner, "Intellectual Property Management Challenges Arising from Pervasive Digitalisation: The Effect of the Digital Transformation on Daily Life," Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data 5, 68, https://www.ictsd.org/sites/default/files/research/ceipi-ictsd_issue_5_final_0.pdf.
25. Alan McQuinn, "Australia's Embarrassing Crackdown on Its Own Information Security," *ITIF Innovation Files,* December 19, 2018, https://itif.org/publications/2018/12/19/australias-embarrassing-crackdown-its-own-information-security.
26. Daniel Castro and Alan McQuinn, "Unlocking Encryption: Information Security and the Rule of Law" (Information Technology and Innovation Foundation, March 2016), https://itif.org/publications/2016/03/14/unlocking-encryption-information-security-and-rule-law.
27. Daniel Castro and Alan McQuinn, "Unlocking Encryption: Information Security and the Rule of Law" (Information Technology and Innovation Foundation, March 2016), http://www2.itif.org/2016-unlocking-encryption.pdf.
28. Reto Hilty, *Big Data: Ownership and Use in the Digital Age, 89.*
29. Wolfgang Kerber, "A New (Intellectual) Property Right For Non-Personal Data? An Economic Analysis," MAGKS Papers on Economics, (2016), 11, 989–998, https://www.uni-marburg.de/fb02/makro/forschung/magkspapers/paper_2016/37-2016_kerber.pdf.
30. Cited in Duch-Brown, Martens, Mueller-Langer, "The Economics of Ownership, Access and Trade In Digital Data"; Andreas Wiebe, *Who Owns Non-Personal Data? Legal Aspects and Challenges Related To The Creation of New 'Industrial Data Rights,'* (2016), Slides presented at the GRUR conference on data ownership, Brussels, Belgium.
31. This definition was developed from several sources, including Erika McCallister, Tim Grance, and Karen Scarfone, "Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)," National Institute of Standards and Technology, 2010, accessed October 24, 2017, http://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800- 122.pdf; The Privacy Act of 1974, 5 U.S.C. § 552a; and "Comments of the Staff of the Bureau of Consumer Protection of the Federal Trade Commission Before the Federal Communications Commission," Federal Trade Commission, May 27, 2016, accessed October 24, 2017, https://www.ftc.gov/system/files/documents/advocacy_documents/comment-staffbureau-consumer-protection-federal-trade-commission-federal-communications-commission/160527fcccomment.pdf.

32. The definitions for "observed information" and "computed information" are similar to the ones the Article 29 Working Group has used for "observed data" and "inferred data." See Article 29 Data Protection Working Party, "Guidelines on the right to data portability," December 13, 2016, revised April 5 2017, 10, https://iapp.org/media/pdf/resource_center/WP29-2017-04-data-portability-guidance.pdf.

33. Wikipedia contributors, "Robert D. Atkinson," Wikipedia, The Free Encyclopedia, accessed September 15, 2018, https://en.wikipedia.org/w/index.php?title=Robert_D._Atkinson&oldid=858806195.

34. "Big Data and Intangible Property Rights – Can IPRs and Rights in Personal Data Exist in Harmony?" Taylor Wessing, March 2017, https://www.taylorwessing.com/download/article-big-data-and-intangible-property-rights.html.

35. "Digital privacy rights require data ownership," *Financial Times*, March 21, 2018, accessed July 20, 2018, https://www.ft.com/content/a00ecf9e-2d03-11e8-a34a-7e7563b0b0f4.

36. Lewis Crofts and Robert McLeod, "MLex Interview: Margrethe Vestager," *MLex*, January 2015, 5, http://mlexmarketinsight.com/wp-content/uploads/2015/01/mlex-interview-vestager-22-01-151.pdf.

37. Herbert Zech, "Information as Property," JIPITEC, (2015), 192, https://www.jipitec.eu/issues/jipitec-6-3-2015/4315/zech%206%20%283%29.pdf.

38. Lee Schafer, "How Much Are Your Online Data Really Worth*?," PHYS.ORG*, April 12, 2018, https://phys.org/news/2018-04-online-worth.html.

39. "Jennifer Ouellette, "Economists Calculate the True Value of Facebook to Its Users In New Study," *Ars Technica,* December 27, 2018, https://arstechnica.com/science/2018/12/economists-calculate-the-true-value-of-facebook-to-its-users-in-new-study/.

40. Duch-Brown, Martens, Mueller-Langer, "The Economics of Ownership, Access and Trade In Digital Data."

41. Geoffrey Manne and Ben Sperry, "The Problems and Perils of Bootstrapping Privacy and Data Into an Antitrust Framework," CPI Antitrust Chronicle, May 2015, 4.

42. Mary Jo Foley, "Microsoft Is Privately Testing 'Bali,' A Way To Give Users Control Of Data Collected About Them," Zdnet, January 3, 2019. https://www.zdnet.com/article/microsoft-is-privately-testing-bali-a-way-to-give-users-control-of-data-collected-about-them/.

43. Rob Atkinson et al., "Manufacturing Digitalization: Extent of Adoption and Recommendations for Increasing Penetration in Korea and the U.S." (Information Technology and Innovation Foundation, August 2018), https://itif.org/publications/2018/08/13/manufacturing-digitalization-extent-adoption-and-recommendations-increasing.

44. Data Policy and Innovation (Unit G.1), Building a European Data Economy, (European Commission, August 2018), https://ec.europa.eu/digital-single-market/en/policies/building-european-data-economy.

45. "Ag Data's Core Principles: The Privacy and Security Principles for Farm Data," Ag Data Transparent, accessed September 18, 2018, https://www.agdatatransparent.com/principles/.

46. Viktor Mayer-Schönberger and Thomas Ramge, "A Big Choice for Big Tech," Foreign Affairs, September 2018, https://www.foreignaffairs.com/articles/world/2018-08-13/big-choice-big-tech.

47. Ibid.

48. Mark MacCarthy, "Data Sharing: A Problematic Idea in Search of a Problem to Solve," CIO, August 20, 2018, https://www.cio.com/article/3301175/regulation/data-sharing-a-problematic-idea-in-search-of-a-problem-to-solve.html.

49. Lisa Kimmel and Janis Kestenbaum, "What's Up With WhatsApp? A Transatlantic View on Privacy and Merger Enforcement in Digital Markets," *Antitrust*, Fall 2014, 53, (citing both the Nielson Holdings/Arbitron and Reuters/Thompson mergers), https://www.crowell.com/files/Whats-Up-With-WhatsApp.pdf.

50. This section is based on an ITIF report: Joe Kennedy, "The Myth of Data Monopoly: Why Antitrust Concerns About Data Are Overblown" (Information Technology and Innovation Foundation, March 2017), https://itif.org/publications/2017/03/06/myth-data-monopoly-why-antitrust-concerns-about-data-are-overblown.

51. For the most comprehensive argument see Maurice E. Stucke and Allen P. Grunes, *Big Data and Competition Policy* (New York: Oxford University Press, 2016).

52. Margrethe Vestager, "Big Data and Competition" (speech before the EDPS-BEUC Conference on Big Data, Brussels, September 29, 2016), http://ec.europa.eu/commission/2014-2019/vestager/announcements/big-data-and-competition_en.

53. European Commission, "Online Platforms and the Digital Single Market Opportunities and Challenges for Europe" (communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, May 25, 2016 COM(2016) 288), 13, http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52016DC0288. See also UK Competition and Markets Authority (CMA), "Online Platforms and the EU Digital Single Market" (written evidence, (OPL0055), CMA, London, October 23, 2015), http://data.parliament.uk/writtenevidence/committeeevidence.svc/evidencedocument/eu-internal-market-subcommittee/online-platforms-and-the-eu-digital-single-market/written/23391.html. "To the extent that such data is of central importance to the offering but inaccessible to competitors, it may confer a form of 'unmatchable advantage,' making it hard for those competitors to compete"; Organization for Economic Co-operation and Development (OECD), Committee for Information, Computer and Communications Policy, "Exploring Data-Driven Innovation as a New Source of Growth: Mapping the Policy Issues Raised by 'Big Data,'" (Paris: OECD, Directorate for Science, Technology and Industry, June 18, 2013), http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=DSTI/ICCP(2012)9/FINAL&docLanguage=En. "Data are a core asset that can create significant competitive advantage and drive innovation, sustainable growth, and development."

54. A McKinsey study estimated that improved use of data could create $3 trillion in additional value within seven industries, including $1.3 trillion in the United States. James Manyika et al., *Open Data: Unlocking Innovation and Performance with Liquid Information* (McKinsey Global Institute, October 2013), http://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx. Another McKinsey study estimated that in 2010, a range of free Internet services increased the social surplus within the United States and the European Union by €120 billion, 85 percent of which went to consumers. IAB Europe and McKinsey, *Consumers Driving the Digital Uptake: The Economic Value of Online Advertising-Based Services for Consumers* (Brussels: IAB Europe, September 2010), https://www.youronlinechoices.com/white_paper_consumers_driving_the_digital_uptake.pdf.

55. Sreedhar Potarazu, "Why Obama's Crack Down on Corporate Mergers Is the Right Prescription for US Health Care," *FoxNews Opinion,* April 16, 2016, http://www.foxnews.com/opinion/2016/04/16/why-obamas-crack-down-on-corporate-mergers-is-right-prescription-for-us-health-care.html.

56. Anja Lambrecht and Catherine E. Tucker, "Can Big Data Protect a Firm From Competition?" *Antitrust Chronicle 1*, no. 12 (January 2017).

57. Ibid., 17.

58. Nestor Duch-Brown, Bertin Martens, and Frank Mueller-Langer, "The Economics of Ownership, Access and Trade In Digital Data" (working paper, JRC Digital Economy, European Commission, Spain, 2016), https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf.

59. "Will Artificial Intelligence Help to Crack Biology?" *The Economist*, January 7, 2017, https://www.economist.com/science-and-technology/2017/01/07/will-artificial-intelligence-help-to-crack-biology.

60. Cade Metz, "Google Just Open Sourced TensorFlow, Its Artificial Intelligence Engine," *Wired*, November 9, 2015, https://www.wired.com/2015/11/google-open-sources-its-artificial-intelligence-engine/.

61. Ibid.

62. Stuart Dredge, "MySpace—What Went Wrong: 'The Site Was a Massive Spaghetti-Ball Mess,'" *The Guardian,* March 6, 2015, https://www.theguardian.com/technology/2015/mar/06/myspace-what-went-wrong-sean-percival-spotify.

63. The European Parliament and of the Council, Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, Article 7, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A31996L0009.

64. "Where's Street Bump Being Used?" accessed September 18, 2018, http://www.streetbump.org/.

65. Daniel Castro and Alan McQuinn, "Cross-Border Data Flows Enable Growth in All Industries" (Information Technology and Innovation Foundation, February 2015), https://itif.org/publications/2015/02/24/cross-border-data-flows-enable-growth-all-industries.

66. Nigel Cory, "Cross-Border Data Flows: Where Are the Barriers, and What Do They Cost" (Information Technology and Innovation Foundation, May 2017), https://itif.org/publications/2017/05/01/cross-border-data-flows-where-are-barriers-and-what-do-they-cost.

67. CEIPI-ICTSD, Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data 5, 87, https://www.ictsd.org/sites/default/files/research/ceipi-ictsd_issue_5_final_0.pdf.

68. European Commission, "Communication from the Commission to the European Parliament, The Council, The European Economic and Social Committee and The Committee of The Regions "Building a European Data Economy," (2017), 8, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2017%3A9%3AFIN.

69. European Commission, "Call for Applications for the Selection of Members of the Expert Group "Business-To-Government (B2G) Data Sharing on Access to and (Re)Use of Private Sector Data for Public Interest Purposes," http://ec.europa.eu/information_society/newsroom/image/document/2018-34/cnect___call_for_applications_b2g_datasharing_38FEAA12-BAF3-5FE2-3894FFBBD501B86D_53961.pdf.

70. For a recent example of this, see the proposed regulations recently issued by the European Commission that limit Internet tracking. Natalia Drozdiak, "EU Proposes New Rules That Could Limit Web Tracking for Ads," *The Wall Street Journal*, January 10, 2017, http://www.wsj.com/articles/eu-proposes-new-rules-that-could-limit-web-tracking-for-ads-1484051998.

ACKNOWLEDGMENTS

The author wishes to thank Daniel Castro and a number of respondents at the George Mason Center for the Protection of Intellectual Property Conference Sixth Annual Fall Conference for providing valuable input to this report. The author also thanks MacKenzie Wardwell for editorial assistance. Any errors or omissions are the author's alone.

ABOUT THE AUTHOR

Robert D. Atkinson is the founder and president of ITIF. Atkinson's books include *Big is Beautiful: Debunking the Myth of Small Business* (MIT, 2018), *Innovation Economics: The Race for Global Advantage* (Yale, 2012), and *The Past and Future of America's Economy: Long Waves of Innovation That Power Cycles of Growth* (Edward Elgar, 2005). Atkinson holds a Ph.D. in city and regional planning from the University of North Carolina, Chapel Hill, and a master's degree in urban and regional planning from the University of Oregon.

ABOUT ITIF

The Information Technology and Innovation Foundation (ITIF) is a nonprofit, nonpartisan research and educational institute focusing on the intersection of technological innovation and public policy. Recognized as the world's leading science and technology think tanks, ITIF's mission is to formulate and promote policy solutions that accelerate innovation and boost productivity to spur growth, opportunity, and progress.

FOR MORE INFORMATION, VISIT US AT WWW.ITIF.ORG.