INFORMING AND REFORMING THE MARKETPLACE OF IDEAS: 
THE PUBLIC-PRIVATE MODEL FOR DATA PRODUCTION 
AND THE FIRST AMENDMENT

Shubha Ghosh*

I. INTRODUCTION

The First Amendment shapes the marketplace of ideas. But this metaphorical marketplace has increasingly been inundated with information. Traditional media like television, radio, and movies; middle-aged media like the Internet; and emerging technologies such as smart phones, peer-to-peer services, and digitally mediated social networking produce new forms of expressive content. These new forms of expression are flush with data—both informative and entertaining—aimed at the new brand of digital citizens and their communities. The marketplace of ideas has become a marketplace of information and data, which is open for those who can penetrate the layers of e-mails, blog posts, tweets, pictures uploaded to Flickr, and Facebook “pokes.” As information and data displace ideas in the metaphorical marketplace, it is no surprise that the First Amendment has taken on new contours, especially with respect to commercial speech, as the United States Supreme Court’s 2011 decision in Sorrell v. IMS Health Inc. illustrates.

This Article presents a critique of what appears to be a new approach to the regulation of commercial speech under the First Amendment. This Article identifies a misdirected normative position regarding how the Supreme Court deals with new technologies for information production and data mining. As Justice Kennedy stated in Sorrell:

The capacity of technology to find and publish personal information, including records required by the government, presents serious and unresolved issues with respect to personal privacy and the dignity it seeks to secure. In considering how to protect those interests, however, the State cannot engage in content-based discrimination to advance its own side of a debate.

* © 2012 Shubha Ghosh. Vilas Research Fellow & Professor of Law, University of Wisconsin Law School.

1 The First Amendment demonstrates the constitutional importance of maintaining a free marketplace of ideas—a marketplace that provides access to “social, political, aesthetic, moral, and other ideas and experiences.” Red Lion Broad. Co. v. FCC, 395 U.S. 367, 390 (1969); see Abrams v. United States, 250 U.S. 616, 630–31 (1919) (Holmes, J., dissenting).


3 Id. at 2681.
Justice Kennedy’s statement demonstrates that the Court is aware of the problems that the new marketplace of information can create for personal privacy and dignity. While the Court is aware of the issues raised by the recent explosion of information, its solution is to let the marketplace regulate itself. Consequently, the Court demonstrated its normative preference for a hands-off approach to the marketplace and struck down the regulations of data mining on First Amendment grounds. Specifically, under the facts of Sorrell, the Court examined Vermont’s attempts to restrict pharmaceutical companies’ use of pharmacy data to directly market to prescribing physicians. While Justices Breyer, Ginsburg, and Kagan would uphold the specific regulation under the First Amendment, their dissent does not fully address the normative grounds for regulating data commercialization consistent with the First Amendment. This Article argues that regulation of the marketplace of information is consistent with freedom of speech and can protect the normative values that support a vibrant marketplace of ideas and the First Amendment.

As applied to the marketplace for information and data, the Supreme Court’s First Amendment jurisprudence—as articulated by the Sorrell majority—emphasizes transparency as an important consequence of commercialization. The Court stated in Sorrell, “The commercial marketplace, like other spheres of our social and cultural life, provides a forum where ideas and information flourish.” But this approach ignores the tension between transparency and commercialization. This Article identifies three compelling normative positions that recognize and address this tension: (i) the classic liberal perspective, (ii) the autonomy perspective, and (iii) the fairness perspective.

Stated briefly, under the liberal perspective, which is perhaps closest to the current Court’s jurisprudence, there is little tension between transparency and

---

4 Id. at 2685.
5 Id. at 2673.
6 Id. at 2671–72 (quoting Edenfeld v. Fane, 507 U.S. 761, 767 (1993)).
8 For an example of the autonomy perspective, see DANIEL J. SOLOVE, THE DIGITAL PERSON: TECHNOLOGY AND PRIVACY IN THE INFORMATION AGE 8–9 (2004) (describing the problem of information privacy as a matter of individual loss of control, especially in the face of a large bureaucracy).
9 For an example of the fairness perspective, see EDITH BROWN WEISS, IN FAIRNESS TO FUTURE GENERATIONS: INTERNATIONAL LAW, COMMON PATRIMONY, AND INTERGENERATIONAL EQUITY 258–59 (1989) (advocating intergenerational equity as a guiding principle in the management of knowledge systems, which includes archiving and preserving data about natural and cultural worlds).
commercialization because a fully autonomous person is one who acts in both the political and market arenas. The autonomy perspective, however, sees potential for conflict as the commercialization of data leads to a confounding of personal and public spaces that can distort meanings and identity. Finally, under the fairness perspective, transparency and commercialization are irreconcilable since market commercialization will lead a corporate culture of commodification to trump other more human and socially desirable values. These three positions are ideal categories, useful in sorting through the various arguments identified in this Article. Each perspective casts light on the policy issues raised by data commercialization and helps conceptualize data commercialization in such a way that it can be organized in an effective and desirable manner.

This Article argues that contemporary debates about the commercialization of data reflect an inherent tension between democratic values of transparency and accountability and the market goal of wealth creation. On the one hand, democratic society demands transparency and its disinfecting effect, which mandates that information and data be open and accessible. On the other hand, wealth creation through markets requires that data and information be processed and transformed, often in ways that distort underlying values and the meanings derived from individual quanta of data and information. This Article explores this tension by demonstrating different ways in which the tension is identified and reconciled. This Article concludes that some form of open-source licensing—which includes protection of personal privacy and autonomy—would be the appropriate business model through which data and information are commercialized.

The commercial use of pharmacy data at issue in *Sorrell* is just one example that implicates the First Amendment, personal privacy, and dignity issues raised by the commercialization of data. Other examples show how values of openness conflict with proprietary and secret uses of information. For example, Wikileaks recently raised issues of what level of transparency could and should exist in an open government. Julian Assange, founder of Wikileaks, advocates for near-complete transparency of government information, which comes into conflict with the needs of security and confidentiality. But the Wikileaks controversy overshadowed more pervasive concerns about the commercialization of data. Social network sites challenge the balance between private and public

---

Highlighting these concerns was a memo from Google, leaked in August 2010, which revealed many of the contemporary dilemmas over the commercialization of data. The memo described various business strategies Google considered that utilized information on consumption patterns obtained from its users of e-mail and search engine services. Google, a trendsetter in market development based on digital technologies and information services, expressed hesitation at the idea of compromising privacy for profit while describing innovative platforms for the exchange of data and advertising to target specific needs of the company’s customer base. The self-described “vision statement” portends the future of consumer-generated and business-processed data as the preeminent commodity of information-based economies.

The issues raised by Google’s vision of data commercialization extend beyond accepted Internet transactions like e-mailing, searching, and social networking. Innovations in the processing of genetic information, particularly as applied to personalized medicine and the possibility of genetic profiling, raise analogous questions regarding the extent to which data commercialization is constrained by social concerns over privacy, individual autonomy, and legal concerns over the scope of property rights. Internet transactions and genetic information raise the same broad question: To what extent can information generated by individuals be the basis for business models focused on the collection, aggregation, and processing of information?

---

11 See James Grimmelmann, Saving Facebook, 94 IOWA L. REV. 1137, 1149–51 (2009) (outlining the contours of privacy protection on Facebook and presumably other social networking sites); see also Jose Antonio Vargas, The Face of Facebook, NEW YORKER, Sept. 20, 2010, at 54–63 (explaining the dynamic between information disclosure and marketing for users of Facebook).


13 Id.

14 Id.

15 Id.


17 See A Sea of Sensors, ECONOMIST, Nov. 6, 2010, at S6–S7 (describing information gathering sensors and their relevance for commercialization, pricing, and marketing); Mark Halper, London: Turning Access into Apps, TIME, Jan. 17, 2011, at 6 (explaining how state-collected information serves as the basis for iPhone applications).
This question, however, extends beyond the realm of new information technologies. For example, one issue that arose during the 2010 U.S. Census was the recording of the relationship status of same-sex couples on the census form. The recognition of marital status marks a progressive turn in equalizing the treatment of same-sex and different-sex relationships. At the same time, advertisers heralded the shift because it allowed for more effective marketing and targeting of advertising dollars to same-sex couples.

Each of these examples—the Google memo, advances in personalized medicine, and census data collection, all of which arise out of very different technological, legal, and social environments—raises a common question: how can data and information be used in a liberal, democratic society that is based on a market economy? This Article addresses these questions as matters of legal entitlements and institutions and as the ways by which society constructs meaning from these questions.

Government policy attempts to navigate the tension between transparency and commercialization when it disseminates data, but often in misdirected ways. For example, the Obama administration created a website, Data.gov, which makes accessible a range of government databases, covering such topics as atmospheric data, employment statistics, and government expenditures. The terms of use for the government site allow the distribution or commercialization of the data, so long as the data is distributed without restrictions on its reuse. Furthermore, the terms of use make a distinction between data and information. Data is defined as “values or sets of values representing a specific concept or concepts.” According to the site, “data become ‘information’ when analyzed and possibly combined with other data in order to extract meaning, and to provide context.” In other words, data is raw or uncooked material. Conversely, information is processed or interpreted data. This distinction between information and data is important in light of the prohibitions on restrictions on reuse of data. Implicitly, the terms of use allow restrictions on reuse of information. By making a distinction between data and information, the government leaves a domain that is open and accessible (data), while allowing the possibility for a proprietary domain where market transactions can be purely private and value can be appropriated (information).

---


23 Id.
The distinction between data and information is an ineffective way through which the conflicting values of privacy, autonomy, and proprietary rights can be reconciled, because the distinction is elusive and perhaps nonexistent. At best, the distinction is one of degree rather than kind. What is missing is a careful consideration of the normative issues raised by the commercialization of data. By limiting the government’s ability to regulate the commercialization of information and data, the Supreme Court in *Sorrell* leaves the matter to the marketplace. Yet this Article demonstrates how regulation of data commercialization can be consistent with the First Amendment through a more careful consideration of the normative goals of liberty, autonomy, and fairness, each of which informs both the First Amendment and the purposes of a marketplace for information and data.

The structure of this Article is as follows: Section II presents an analysis of the *Sorrell* decision, placing it in the context of First Amendment jurisprudence. It also includes a brief history of information and data commercialization. Section III builds on economic realities by turning to the legal structure of data markets, which are rooted in intellectual property and other information-regulating laws such as privacy and security. Section IV presents a critique of what is described as the emerging commercial speech paradigm. Section V summarizes and concludes that some form of open-source licensing, with adequate protections for personal privacy and autonomy, is the best model for the commercialization of data.

II. *SORRELL* AND THE MARKETPLACE OF IDEAS, INFORMATION, AND DATA

The Supreme Court’s 2011 decision in *Sorrell* is a touchstone for understanding the normative issues raised by the commercialization of data. Unfortunately, both the majority and dissenting opinions failed to adequately address the broader normative and policy concerns implicated by the case—such as the risks to privacy and dignity accompanying the commercialization of data. Instead, the opinions framed the rich facts of the dispute within the narrow confines of First Amendment jurisprudence. Nonetheless, an examination of the opinion is a critical starting point for assessing the economic, social, and political values raised by data commercialization. This section presents that examination, while subsequent sections build upon this analysis of the opinion to construct a deeper understanding of the jurisprudence of commercialized data.

At issue in *Sorrell* were state regulations that restricted the ability of individuals to use pharmaceutical data to target prescribing doctors with advertising and promotions.\(^{24}\) This practice, called “detailing,” is used by

---

\(^{24}\) The statute at issue is Vermont’s Prescription Confidentiality Law. *Vt. Stat. Ann.* tit. 18, § 4631(d) (2009). The provisions relevant to this case state that “a health insurer, a self-insured employer, an electronic transmission intermediary, a pharmacy, or other similar entity shall not sell . . . regulated records which include prescription information containing prescriber identifiable data . . . unless the prescriber consents.” *Id.* In addition, “[p]harmaceutical manufacturers and pharmaceutical marketers shall not use prescriber-
pharmaceutical companies with patent-protected pharmaceuticals whose sales representatives contact physicians directly to promote their products. Typically, private companies that aggregate and package data for purchase by the pharmaceutical companies—referred to as data miners—collect the data. Some states enacted regulations that prohibit this practice by preventing pharmacies and insurers from selling pharmaceutical data, as well as preventing pharmacies, insurers, and pharmaceutical companies and marketers from using such data for marketing purposes.

The states enacted these restrictions on detailing to contain the costs of health care, particularly prescription drugs. Large pharmaceutical companies were more likely to use detailing practices than companies that provided cheaper, generic alternatives. Furthermore, the states sought to protect the privacy of physicians whose information was contained in the pharmaceutical data. Relevant privacy statutes, such as HIPAA, protected the patient’s privacy but not the doctor’s. The prohibitions on detailing were meant to protect doctors from intrusive advertising and marketing. Claiming violations of their First Amendment rights, data miners and pharmaceutical companies challenged the statutes. The Maine and New Hampshire statutes were found constitutional by the First Circuit in *IMS Health Inc. v. Mills* and *IMS Health Inc. v Ayotte*, respectively. But the Second Circuit struck down Vermont’s statutes in *IMS Health Care v. Sorrell*. This circuit split led the Supreme Court to grant certiorari in *Sorrell*.

In a six-to-three decision, the Supreme Court affirmed the Second Circuit and concluded that Vermont’s prohibitions on the sale and use of data were content- and speaker-based restrictions on speech. The dissent would have upheld the regulations as “a lawful governmental effort to regulate a commercial enterprise” and justified the harm to any First Amendment interests as proportionate to the “furtherance of legitimate regulatory objectives.” The majority and dissent differ in (a) their respective conceptions of the speech at issue, (b) the nature of the identifiable information for marketing or promoting a prescription drug unless the prescriber consents.”

---

26 Id. at 2660.
27 Id. at 2661 (explaining “counter detailing” statutes that exist in various other states).
28 Id. at 2668.
29 Id.
30 Id. at 2684 (Breyer, J., dissenting); see also Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936.
31 Id. at 2668 (majority opinion).
32 616 F.3d 7, 48–49 (1st Cir. 2010).
33 550 F.3d 42, 105–06 (1st Cir. 2008).
34 630 F.3d 263, 281–82 (2d Cir. 2010).
35 *Sorrell*, 131 S. Ct. at 2659.
36 Id. at 2673 (Breyer, J., dissenting).
regulation, and (c) the effect of the regulation on First Amendment values. As argued below, neither the majority nor dissent addresses the core economic, social, and political values raised by data commercialization beyond the narrow framing of the issues as a matter of freedom of speech. The majority blurs the line between commercial regulation that has an impact on speech and direct regulation of speech. In so doing, the Court moves away from its traditional intermediate scrutiny standard for commercial speech, effectively abandoning the substantial interest of protecting against intrusions on privacy and dignity. But before addressing these broader issues, the next subsection will analyze the critical differences between the majority and the dissenting opinions.

A. The Nature of the Speech at Issue

Since the regulation at issue targets the sale and marketing of data, the speech at issue arguably falls into the category of commercial speech, which is protected under a lower standard of review than core categories such as political speech. A government restriction on commercial speech is reviewed under an intermediate standard of review, which requires the government to show that “the statute directly advances a substantial governmental interest and that the measure is drawn to achieve that interest.” The dissent found that the Vermont statute met this requirement because it attempted to balance prescribing doctors’ communicative channels that the state had deemed biased toward large pharmaceutical companies. The government’s goals to reduce health care costs and protect citizens by protecting doctors from intrusive advertising satisfied the requirements of intermediate scrutiny.

The majority, however, construed the speech at issue differently. Although the opinion recognized that commercial speech was at issue, the majority concluded that the government failed to meet its burden under the First Amendment. Justice Kennedy, writing for six justices, went beyond the intermediate standard of review, reasoning that strict scrutiny should apply. Initially describing the speech at issue as “speech in aid of pharmaceutical marketing,” Justice Kennedy concluded that the state regulation was based on the content of speech and the identity of the speaker.

First, the majority deemed the restriction content-based because it targeted speech that promoted the marketing of name-brand pharmaceuticals, ostensibly to promote or subsidize speech used to market generics and low-cost substitutes.

37 Id. at 2667–68 (majority opinion).
38 Id. at 2580–84 (Breyer, J., dissenting).
39 Id. at 2667–72 (majority opinion).
40 Id. at 2668–72.
41 Id. at 2663–65.
42 Id. at 2659.
43 Id. at 2663–65.
44 Id. at 2663.
According to the majority, the state had chosen sides in the advertising marketplace by attempting to shut down one channel of communication in an attempt to remove the advantage that large pharmaceutical companies had in acquiring and analyzing market data.\(^{45}\) The regulation was not a broad-based commercial regulation that had some incidental effect on speech,\(^{46}\) nor was it targeted at specific types of commercial speech, such as attorney advertising, across the board.\(^{47}\) Instead, the restriction on detailing favored one advertiser at the expense of another by prohibiting a specific marketing practice.\(^{48}\) This regulatory strategy made the statute a content-based restriction on speech, subject to strict scrutiny. The Court required the state to show a compelling interest, not just a substantial one, and a close nexus between that interest and the means chosen. According to the majority, Vermont had failed to do so.

Second, the majority found that the restriction was based on the identity of the speaker.\(^{49}\) The targets of the restriction were pharmaceutical companies that used the data purchased from detailers to communicate with doctors to affect their prescription practices.\(^{50}\) Although Vermont, in a seemingly last-ditch effort, contended that the restriction was aimed at the sale and use by any entity that obtained the data from the pharmacies, the majority was not convinced.\(^{51}\) But even if the Court accepted the State’s final contention, the majority opinion emphasized exceptions in the Vermont statute that allowed pharmacies to sell the data to universities or other entities for research purposes.\(^{52}\) The statute, the majority concluded, was aimed not only at specific content, but also at specific speakers.\(^{53}\) Based on the foregoing, the Court concluded that strict scrutiny would apply on either ground.

From its start, Justice Kennedy’s opinion narrowly construed the speech as “speech in aid of pharmaceutical marketing.”\(^{54}\) The majority rejected the argument—that the speech at issue was a mere commodity that was bought and sold rather than communicative expression.\(^{55}\)
Court demonstrated its rejection of this characterization when it stated, “Facts, after all, are the beginning point of much of the speech that is most essential to advance human knowledge and to conduct human affairs. There is thus a strong argument that prescriber-identifying information is speech for First Amendment purposes.”

The pieces of data at issue were not mere commodities that could be regulated by the state but valuable components of inputs to decision-making and the crystallization of knowledge and opinions. In other words, the Court made it clear that in a marketplace of ideas, citizens making decisions about pharmaceuticals require access to such data. While states could have an interest in regulating such data as speech if it were false, the majority found the speech to be useful in physician and patient decision-making. Therefore, the state could not simply restrict the speech because of its undesirable results, such as intrusiveness into privacy or higher health care costs.

The majority and dissent start from two different characterizations of speech and reach diametrically opposed conclusions. If data were viewed as pure commercial speech, the communications at issue under the regulation—the transfer of data, its production, and its use for advertising to doctors—would be subject to state regulation with slight protection under the First Amendment. If viewed as a very narrow communication about the merits of certain products to certain consumers, the speech in question could not be regulated because the state had singled it out based on its content and the identity of the speaker. Both the dissent and majority agree on the value of the marketplace of ideas, but they differ on how that marketplace operates when dissemination of information is its object.

B. Competing Visions of the Nature of the Regulation

The majority and dissent also diverged in their characterizations of the regulation. As discussed above, the majority viewed the regulation as based on the content of the speech and the identity of the speaker. The dissent, by contrast, saw the regulation as part of a comprehensive scheme to regulate pharmaceuticals and protect the interests of patients. Not surprisingly, the two contrasting depictions led to opposing assessments of the First Amendment claim.

As the majority acknowledged, all government regulation has some effect on speech. To take the majority’s own example, laws against employment discrimination affect what and how messages are communicated in advertisements

56 Id. at 2667.
57 Id. at 2671–72.
58 See supra note 24 for relevant language from the Vermont statute. The speakers being targeted, according to the majority, are the pharmaceutical manufacturers and marketers, and the content being singled out is marketing or promoting a prescription drug. Sorrell, 131 S. Ct. at 2665–67.
59 Sorrell, 131 S. Ct. at 2673–75 (Breyer, J., dissenting).
60 Id. at 2664–65 (majority opinion).
for jobs and in everyday office conversation. But such effects do not transgress the First Amendment because they are deemed incidental—much like restrictions on sexual harassment. According to the majority, the Vermont statute directly regulated speech by restricting use of certain data that enabled communication between marketers and physicians.\(^{61}\) Regulation of speech was central to the Vermont statutory scheme and the goals of protecting privacy and controlling health care costs were deemed incidental.\(^{62}\) The dissent, on the other hand, saw the effects on speech as incidental to the legitimate goals of the state in regulating health care and the pharmaceutical industry.\(^{63}\)

One argument critical to the dissent’s position is that the prescription data at issue exists because of government regulation.\(^{64}\) Federal food and drug regulation requires a prescription for certain medications.\(^{65}\) These regulations require prescriptions to be of a particular form and to be processed in a particular way.\(^{66}\) Absent this requirement, the physician-patient relationship largely would be a matter of pure contract. Physicians might recommend certain drugs as treatment and might even dispense them directly. The structure of pharmaceutical distribution would be very different absent government regulation. Since the prescription data itself is a creature of regulation, the dissent concluded that the government has wide latitude in regulating such data.\(^{67}\) This characterization of the data at issue contrasts with the majority’s view that pharmaceutical data is private data because it is in the hands of a private entity, the pharmacies.\(^{68}\) By emphasizing the private possession of the data, the majority distinguished Supreme Court precedent regarding regulation of criminal files in the possession of the state police, access to which can be restricted without offense to the First Amendment.\(^{69}\)

The majority’s view of the data emphasizes the possessory interest and ignores the governmental regulations that effectively led to the creation of the data. In short, the majority’s analysis pays short shrift to the regulatory environment, and reduces it to a private property interest with ownership based on possession. By contrast, the dissent correctly recognizes the public nature of the pharmaceutical data, but does not adequately address the ownership issue.

The dissent implicitly and incorrectly assumed, however, that because government regulation created the pharmaceutical data, the government owns the

\(^{61}\) Id. at 2663–65.
\(^{62}\) Id. at 2668–69.
\(^{63}\) Id. at 2685 (Breyer, J., dissenting).
\(^{64}\) Id. at 2674–76.
\(^{65}\) Id. at 2675–76, 2682–83.
\(^{67}\) See Sorrell, 131 S. Ct. at 2676–77 (Breyer, J., dissenting).
\(^{68}\) Id. at 2665–66 (majority opinion).
\(^{69}\) Id. (distinguishing Los Angeles Police Dept. v. United Reporting Publishing Corp., 528 U.S. 32, 40 (1999)).
data and can do what it wants with it. That broad conclusion cannot be correct, as a simple hypothetical will show. If the government chose, pursuant to the dissent’s reasoning, to publicize all the data that it created and purportedly owns, then one does not need much imagination to predict the public outcry. The government’s regulation of the data is not pursuant to its sole ownership of the data. Rather, the government’s regulation balances the interests of the many parties that have some stake in the data: the patients, the physicians, and the pharmacists. The majority ignores these disparate interests by treating data as a private property interest.

Both the majority and the dissent characterized the data too simplistically—one overemphasizing its private nature, the other its public roots—in reaching their conclusions about the Vermont regulation. What is missing is a more detailed understanding of the life of data and the intricate ownership and regulatory structure surrounding its creation and dissemination. Section III of this Article provides an alternative approach, which becomes even more necessary after considering how the majority and dissent in *Sorrell* illustrate the problems that arise from not fully understanding the marketplace for data and information.

C. The Marketplace of Ideas

The previous section pointed to some of the normative deficiencies in the majority and dissenting opinions. Nowhere is this more obvious than in the majority and dissent’s contrasting views of the marketplace of ideas, which in turn inform their respective understandings of the First Amendment. Critics of First Amendment jurisprudence have pointed to the inadequacy of the marketplace of ideas as a paradigm for adjudicating free speech rights. The marketplace metaphor assumes some form of free, commodified exchange—a “free trade in ideas.” Critics point to other values that a marketplace metaphor ignores, such as the values of community, pluralism, and communication—captured by the phrase “free play of ideas,” which de-emphasizes the potentially misleading use of the word “trade.” Nonetheless, the majority and dissent in *Sorrell* both rely on the marketplace of ideas, albeit with different implications. Consequently, this section discusses the concept of a marketplace of ideas without ostensibly engaging with its many critics.

The majority views the marketplace of ideas in laissez-faire terms, meaning that an unregulated market leads to the best results in terms of disseminating information to marketplace participants. Under this view, any attempt by the government to favor one side or to tilt the marketplace in one direction would be

---

70 See id. at 2677 (Breyer, J., dissenting).
73 See *Sorrell*, 131 S. Ct. at 2670–72.
inconsistent with the First Amendment. Judicial suspicion toward content-based regulation, as well as regulation based on the identity of the speaker, stems from distaste for the government attempting to intervene into the free trade of ideas made possible by the marketplace.

In *Sorrell*, the majority struck down the Vermont statute precisely because of its particular view of the marketplace, and yet the majority attempted to tilt it in favor of some speakers at the expense of others.\(^74\) Although the State expressed concern that large pharmaceutical companies could gain access to physicians while smaller generic companies could not, the majority attributed this difference in access to the persuasiveness of the pharmaceutical companies.\(^75\) Under this approach, the First Amendment prohibits the government from punishing or prohibiting this persuasive success simply because the government disagrees with the result unless there is a strong justification. Doctors, as the majority pointed out, could simply ignore the marketers,\(^76\) and health care costs could be contained through means that do not interfere with speech.\(^77\) In fact, communication about products facilitates efficient and accurate pricing.\(^78\) Neither protection of physician privacy nor health care cost containment justified Vermont’s restriction. Thus, the state cannot regulate price through restricting advertising.

In all respects, the majority adopted a laissez-faire view of the marketplace of ideas. Unlike the majority, the dissent concluded that the marketplace of ideas fails to be a perfectly functioning market. In other words, the dissent concluded that at the core of the First Amendment is the notion that the government must intervene to protect against typical market deficiencies like concentration of resources and unequal access. As the dissent stated:

> [Court precedent] reflects the need to ensure that the First Amendment protects the “marketplace of ideas,” thereby facilitating the democratic creation of sound government policies without improperly hampering the ability of government to introduce an agenda, to implement its policies, and to favor them to the exclusion of contrary policies.\(^79\)

The Vermont statute, according to the dissent, is one way by which parity can be established in the pharmaceutical market. Because of government regulation, a valuable resource in the form of pharmaceutical data is created. Enterprising and well-heeled companies have the financial means to extract this resource and use it in communicating with physicians to better advertise and sell their products. Statutes such as the one enacted by Vermont limit the advantage enjoyed by

\(^74\) Id. at 2671.
\(^75\) Id. at 2670.
\(^76\) Id. at 2669.
\(^77\) Id. at 2670.
\(^78\) Id. at 2671.
\(^79\) Id. at 2679 (Breyer, J., dissenting).
pharmaceutical companies and permit generics to compete more effectively. 80 Specifically, the Vermont statute prevented "companies from shaping a commercial message they believe maximally effective." 81 Instead of undercutting the marketplace of ideas, the Vermont statute regulated the market for goods and services. 82 Therefore, the dissent concluded that the statute is consistent with the First Amendment. 83

Sorrell is the seminal case for the constitutional standard by which to gauge regulation of the market for information and data. Both the majority and dissent construed this marketplace in terms of the marketplace of ideas—the workhorse of First Amendment analysis. But as the previous section established, both the majority and dissent are deficient in their normative assessment of the regulation and the free speech values at issue. Although the two sides are divided in their assessments of the marketplace of ideas, notably, both failed to see beyond the concept of a marketplace as the lens for understanding how information and data are generated, disseminated, and regulated.

This focus on the marketplace of ideas as the sole institutional context within which to gauge state regulation is underscored by the majority’s understanding of the treatment of commercial speech under the First Amendment. As the dissent pointed out, the Court has traditionally been relatively deferential to regulation of commercial speech—such as commercial advertising—by subjecting such regulation to only intermediate scrutiny. 84 In its 1976 decision in Virginia State Board of Pharmacy v. Virginia Citizens Consumers Council, Inc, 85 the Court stated that there were “commonsense differences” between commercial speech and other categories of speech, and therefore commercial speech warranted a different level of protection. 86 This explains why the dissent in Sorrell took issue with the skeptical majority’s use of strict scrutiny.

The majority in Sorrell blurred the distinction between commercial and non-commercial speech. 87 Citing language from the Court’s 1977 decision striking down a blanket ban on attorney advertising, the majority reminds us that “[a] consumer’s concern for the free flow of commercial speech often may be far

80 Id. at 2677 (“[T]his Court has never found that the First Amendment prohibits the government from restricting the use of information gathered pursuant to a regulatory mandate.”).
81 Id. at 2681.
82 Id. at 2674 (“[O]ur cases make clear that the First Amendment offers considerably less protection to the maintenance of a free marketplace for goods and services.”).
83 Id. at 2685.
86 Id. at 771 n.24.
keener than his concern for urgent political dialogue."88 The Sorrell majority underscored this point as follows: “That reality has great relevance in the fields of medicine and public health, where information can save lives.”89 The reality the Court referred to is the need for free flow of information, a need the Court saw as coequal in the commercial and political contexts.

There are values relevant to the commercialization of data that are not encompassed by the marketplace of ideas.90 These values, which are examined through the normative lenses of the liberal perspective, the autonomy perspective, and the fairness perspective, can be realized through a combination of the marketplace and of regulation. The next section will flesh out a richer legal and normative structure of the marketplace for information and data. Section IV examines other examples of such marketplaces and returns to the case of Sorrell in light of this richer legal and normative structure.

III. THE LEGAL AND NORMATIVE STRUCTURE OF INFORMATION MARKETS

This section elaborates upon the Court’s failure to adequately confront the normative issues raised by the commercialization of data in its Sorrell decision. Both the majority and dissent framed the regulation of data mining as a restriction on the marketplace of ideas. But the marketplace of information and data is far richer than the metaphorical marketplace of ideas. One root of the problem is that litigants and commentators often fail to fully describe the legal structure of information markets and their normative implications. The following discussion explicates both the legal and normative structure of information markets and provides broader context for the controversy in Sorrell.

The legal structure governing data affects how data is created, distributed, and consumed. The legal structure governing data can be divided into two sets of issues: ownership and data transactions. Intellectual property law grounds the ownership issues, particularly copyright, trade secret, and patent. The structure of ownership determines who can license and transfer data and what types of data can be subject to commercialization. Privacy law and norms, as well as the norms of security, are legal structures that govern permissible data transactions, such as the selling of personal data to advertisers or the collection of data for the purposes of profiling.

State creation and ownership of data is an important piece of the legal foundations for data governance—Data.gov is just one example of the governmental role in creating, disseminating, and consuming data. In this Article,

89 Id.
the role of the state will be considered in the discussion of ownership, specifically in the analysis of open records acts, which introduce another layer to the relationship between data and information. As this section and the case examples of Section IV demonstrate, the government is a key player in such situations as illustrated by the commercialization of real estate and census data.

Understanding the legal rules and institutions governing data is a step toward assessing data commercialization against a normative framework. As explained in Section I of this Article, there are three normative perspectives one can take toward the practices of data commercialization: (1) the liberal perspective, (2) the autonomy perspective, and (3) the fairness perspective. Each perspective addresses the twin issues of transparency and commercialization in the management and use of data.

Under the liberal perspective, transparency and commercialization are not incompatible because the market realm provides a form of transparency that recognizes the freedom of individuals. For example, the use of census data about same-sex couples by advertisers is consistent with the liberal perspective. Such a use recognizes the economic value and autonomy of same-sex couples, which can be as important a force as political or legal recognition. The liberal perspective gives equal weight to politics and markets in respecting the rights of individuals.

The autonomy perspective, on the other hand, recognizes that different realms provide different weight to aspects of individual autonomy. Being able to buy and sell is less of an exercise of autonomy than being able to vote or legally marry. Therefore, the legal regime needs to recognize these differences and treat them accordingly. From the autonomy perspective, the census example presented in the previous paragraph is a potential betrayal to an individual’s autonomy, allowing market recognition of same-sex couples while denying political or legal recognition. Legal institutions should respond to this discrepancy by giving full weight to the autonomy of individuals.

Finally, the fairness perspective considers the differing bargaining powers of actors in the political and market realm and scrutinizes the commercialization of data from a distributive justice perspective. For example, the fairness perspective would treat the collection and use of census data as a potential exploitation of same-sex couples and would support a ban on the collection or use of the data. This approach contrasts with that of both the liberal and the autonomy perspectives, which would both allow the collection and use of the data. The liberal perspective would view the use of the data as a legitimation of same-sex couples, perhaps even a step toward formal political and legal recognition. The autonomy perspective would be skeptical of it but would demand equal treatment of same-sex couples in the economic, legal, and political realms.

This section presents and assesses the legal structure of data commercialization in light of this normative framework. It shows that the actual legal structure governing data commercialization represents a compromise among these three normative perspectives. In this way, data commercialization is made possible within competing normative perspectives. The analysis presented in this section will provide the foundation for the licensing model for data presented in
Section IV, which is an attempt to reconcile the competing normative forces by acknowledging them more completely.

A. Ownership and Data Commercialization

To what extent can data be owned? How does ownership in data relate to ownership in information? If data and information can be owned, are there restrictions on how each can be sold or shared? These are the questions addressed in the rest of this subsection, which discusses examples drawn from two areas: intellectual property law and data ownership and commercialization by governmental entities.

1. Intellectual Property Regimes

Four bodies of intellectual property law govern the ownership of data. They are discussed in order of their doctrinal and normative complexity: (a) copyright, (b) patent, (c) trade secret, and (d) common law misappropriation. This discussion will conclude with a short analysis of overarching First Amendment principles.

(a) Copyright

The principal United States case directly addressing the question of ownership of data is the 1991 Supreme Court decision in *Feist Publications, Inc. v. Rural Telephone Service Co.* 91 At issue in *Feist* was the copying of entries in a telephone directory compiled and issued by the Rural Telephone Service Company in Kansas. 92 The company issued separate telephone directories for each city in Kansas. Feist took the separate directories for several of the cities and put together a meta-directory that allowed individuals to search the entire state. 93

In ruling that Feist had not infringed the company’s copyrights on the directories, the Supreme Court emphasized two legal principles. First, one requirement for copyright protection is originality. This means that the purported author of the work was the original source for the protected material, and that the author imbued the material with her own creativity. 94 Second, with respect to compilations of data or facts like a telephone directory, copyright protection does not extend to the facts or data compiled because they do not originate from the author. 95 Instead, copyright protection extends to the creative selection, arrangement, and coordination of the facts and data by the author making the

---

92 Id. at 342–44.
93 Id. at 342–43. In 1991, telephone number inquiries were conducted by manual and visual searches, as opposed to digital searches.
94 See id. at 348–49.
95 See id. at 349–51.
compilation. The Rural Telephone Company was deemed not to have a copyright for an alphabetical arrangement of names and other data in a phone directory, and Feist was free to copy that list of names and other data. 96 The Court emphasized that copyright protection for a database was thin, extending only to the particular selection, arrangement, and coordination of the facts and data, creatively chosen by the author.97

The Feist decision conflated the terms facts and data. The Court referred to facts and occasionally referred to data.98 The Court’s rationale, however, applied equally to both facts and data, whatever the distinction. As the etymology of the word suggests, data is “given” by some source other than the person compiling data. Data exists presumably prior to the creation of the database and independent of the compiler.99 Therefore, data cannot be copyrighted because the first criterion for copyright protection is not met: data does not originate from the author.

One still may ponder the relationship between facts and data. Data.gov defines data as values assigned to certain concepts, with the implication that these values are quantitative. Facts, however, can be qualitative as well as quantitative. To take an antiquated example—one safely in the public domain—the fact that it took Phileas Fogg seventy-nine days to circumnavigate the Earth in Around the World in Eighty Days100 contains both qualitative and quantitative facts. The quantitative fact (seventy-nine days) would be data according to the working definition of data. This distinction may not be satisfactory because “values” can be qualitative as well. For this reason, this Article will leave others to make meaningful distinctions between facts and data. As such, this Article will use the terms “data” and “facts” synonymously, particularly when discussing copyright law.

The example of Phileas Fogg illustrates a more important point than the semantic distinction between facts and data—the possibility that data could be fictional. Since originality for copyright purposes requires creativity, fictional data raises interesting possibilities for copyright protection. Copyright doctrine would treat the datum, “Phileas Fogg took seventy-nine days to circumnavigate the Earth,” in a number of ways.101 First, this datum encapsulates an expression, and as an expression the statement can obtain some degree of copyright protection when fixed in a longer work, whether a paragraph like this one or a novel like Jules Verne’s.

---

96 Id. at 362–64.
97 Id.
98 See id. at 345 (referring to raw data as “wholly factual information not accompanied by any original written expression”).
99 Id. at 340.
An additional requirement for copyright protection is that the expression be a recognized work of authorship, such as the compilation at issue in *Feist*.102 Other recognized works of authorship include literary works, musical works, and audiovisual works. A single sentence can constitute a literary work, although the U.S. Copyright Office by regulation has excluded short phrases—such as advertising jingles—from copyright protection to draw a distinction between copyright and trademark laws.103 A longer work containing the sentence, like this paragraph or a novel, would be a literary work, and the copying of the sentence would provide evidence of substantial similarity, the legal test for copyright infringement.104 In this way, copyright protection extends to the ways in which data is expressed.

Second, copyright law extends only to the expression, and not to the idea being expressed.105 The data contained in a work of authorship would not be subject to copyright protection, as the Supreme Court made clear in *Feist*.106 Therefore, anyone can report that Phileas Fogg went around the world in seventy-nine days. As discussed above and under *Feist*, an individual is free to use the data but not the original selection, arrangement, and coordination of data—in other words, the way in which the data is expressed through a database.107 The analogy would be to letters of the alphabet and a novel, or a note in a musical scale and a score. Copyright protection in a work does not extend to building-block components of the work, such as a letter, a note, or data.

Third, a challenging question arises out of the treatment of created or constructed data. In the case of the Phileas Fogg example, Jules Verne initially created the datum in a book that is now in the public domain. To what extent would copyright protection extend to a fictional piece of data? At tension here are the principles of creativity, the lack of copyright protection for data, and the definition of work of authorship. Copyright is meant to promote creativity by giving a time-limited protection to creative works for the purposes of management by the author and any eventual copyright owner.108 This principle favors the protection of created data. On the other hand, data are meant to be openly

---


104 U.S. COPYRIGHT OFFICE, supra note 102, at 3; ROBERT C. OSTERBERG & ERIC C. OSTERBERG, SUBSTANTIAL SIMILARITY IN COPYRIGHT LAW § 1.1 (2012).


107 See id. at 347–48.

accessible and standing alone may not constitute a work of authorship. Add to this conundrum the observation that copyright’s directive is to promote progress in science, an old-fashioned word for knowledge. By recognizing copyright protection for fictional data but not for nonfictional data, would not copyright law be promoting the creation of lies, falsehoods, and misconceptions? This last point is an intriguing one, but perhaps can be addressed by noting that value judgments and judgments of quality are not the domain of copyright law. Copyright law may protect falsehoods, but it is the bailiwick of other areas of law to police them. More troubling is the tension between creativity and access to data.

Case law subsequent to Feist has addressed the issue of constructed or fictional data, particularly with respect to the creation of items like stock indices or sports statistics. In such cases, lower courts have made a distinction between data that would constitute a building block and data that would not constitute a building block. The former cannot be subject to copyright protection to protect accessibility, and the latter can to recognize creativity. The problem, of course, is determining when a piece of data is a building block. On this question, courts are not very helpful and tend to focus on the alterations made by the secondary user of the data who created an index or sports statistic based on the data. The analysis is analogous to the determination of what types of selection, arrangement, and coordination would constitute creativity. Another way to think about this analysis, within the language of this Article, is the distinction between data and information. Courts that purport to protect non–building block data are effectively protecting information that is a creative interpretation of data. Stock indices and sports statistics are examples of how data is processed and analyzed to produce another work that represents an interpretation of the underlying data. Courts, therefore, are not protecting non–building block data, but information, which represents a way of processing the underlying data, like in a database.

---


110 See, e.g., Belcher v. Tarbox, 486 F.2d 1087, 1088 (9th Cir. 1973) (“There is nothing in the Copyright Act to suggest that the courts are to pass upon the truth or falsity, the soundness or unsoundness, of the views embodied in a copyrighted work.”).

111 See Kregos v. Associated Press, 937 F.2d 700, 705 (2d Cir. 1991) (holding the compiling of facts, set apart from the discovering of facts, can be patentable).

112 Id. at 705–06 (discussing the extent to which selection of data must be sufficiently creative to trigger copyright protection).

113 Id. at 704.

114 See supra notes 16–17 and accompanying text.

115 Cf. CDN, Inc. v. Kapes, 197 F.3d 1256, 1260–61 (9th Cir. 1999) (involving price list of collector’s coins).

The approach under *Feist* contrasts with that adopted by the European Database Directive (Directive) in 1996. The European Union issued the Directive in response to concerns by members of the database industry that the thin copyright protection under *Feist* was not adequate in protecting the economic interests of legal and financial database owners. The Directive extends protection for fifteen years to creators of databases that are the product of “substantial investment in obtaining, verifying, or presenting the contents of the database.” The Directive gives database owners the right to prevent imitation or alteration of the database and extraction of the data from the database.

The Directive will be discussed in more detail in Section IV of this Article. At this stage, it is important to identify the differences between the Directive and the approach taken in *Feist*. First, the Directive extends protection to the underlying data itself, to the extent the data is extracted from the database. Second, the Directive bases protection on the creator’s “substantial investment” rather than on a standard of creativity. Arguably, an alphabetic listing of names, such as that in *Feist*, would be protected if the compilation of such a list required substantial investment of time and money. Under *Feist*, copyright law would never protect a mere alphabetic compilation. Therefore, the Directive expands the scope of database protection and rights of the database creator. The United States considered similar legislation in the 1990s, but those seeking to protect access to data blocked passage of analogous laws.

Copyright law has several implications for the commercialization of data. Under the foundational principle of copyright, data would not be subject to copyright protection. Birthdays, nicknames, cities of birth, occupations, names of friends, favorite colors, favorite musical groups, favorite books, and the myriad other pieces of data that are stored on websites or in nondigital form would constitute unoriginal items not subject to copyright, meaning they can be copied by anyone without fear of infringement. Expressions of any length, like posts or comments, would constitute copyrightable subject matter owned by the author of the expression. But these posts and comments can be gleaned, and any data within them can be extracted without the risk of running afoul of copyright. Furthermore, many of these expressions may lose copyright protection because the expression and the idea expressed merge so that one cannot take the data without also taking the expression. For example, if a post says “I partied all night yesterday,” the data contained in the sentence cannot be separated from the sentence itself. Therefore,

---

118 See id. pmbl, para. 2.
119 Id. arts. 7, 10.
120 Id. art. 7.
121 Id.
122 Id.
under the merger doctrine, the sentence itself loses copyright protection.\textsuperscript{124} For these reasons, data is free of copyright protection and ready for anyone to commercialize.

The freedom of access to data under copyright law limits the use of copyright law as a means of extracting commercial value from it. If anyone can copy data, then no one would be willing to pay for data. The value comes from the construction of information and databases from the underlying data. What copyright law permits is an entrepreneur to process raw data to create an interpretative layer in the form of statistics or other composite information that synthesizes or analyzes data into a valuable, and hence marketable, form. Such information could include summary statistics or reportable information that provides value for those unable to aggregate and synthesize freely available data. In addition, value can be created through databases, which entail the compilation of data in creative ways. Databases do not involve the interpretation of data, necessarily, but do involve the packaging of data in ways that allow users to form their own interpretation and extract the data.\textsuperscript{125} Commercialization of data very often involves the marketing of personal data in databases that serve to profile and target potential consumers of advertising and new products.

An interesting question arising from copyright law is what constitutes a database. Data extracted from, for example, a social networking site and arranged and coordinated creatively can constitute a database. But what about the website itself? To what extent is all of Facebook a database? These questions have not been directly addressed, but they acknowledge the existence of meta-databases or databases that consist of other databases.\textsuperscript{126} The layers of complexity made possible by meta-databases may undermine the underlying principle of access to data since at some point it may be difficult, if not impossible, to separate the copyright-protected database from the freely accessible data.\textsuperscript{127} In this way, the owner of the meta-database may end up effectively obtaining ownership of the data and limiting access.\textsuperscript{128} Further complicating the analysis is the number of ownership interests that might constitute the meta-database. Individual users may be responsible for owning portions of the database that they created while the creator of the website itself will stake a claim in the aggregate of the databases. Conflicting and competing ownership interests may mitigate the ability of the...

\textsuperscript{124} See, e.g., N.Y. Mercantile Exch., Inc. v. Intercontinental Exch., Inc., 497 F.3d 109, 116–17 (2007) (denying copyright protection to expressions of ideas that can only be expressed in a very limited number of ways under the merger doctrine, such that the expression and the idea are so intertwined that they “merge” and are both ineligible for protection).

\textsuperscript{125} See Reichman & Samuelson, supra note 123, at 66–67.

\textsuperscript{126} See id. at 114–15.

\textsuperscript{127} See id. at 103–04.

\textsuperscript{128} See id. at 105–06.
meta-database owner to restrict access and thereby continue to keep data free from copyright protection.  

(b) Patent

The manipulation, extraction, and interpretation of data are the primary vehicles of extracting value through data commercialization. Copyright law protects the expressive uses of data through the creation of information and databases. But copyright law would not protect innovative ways of processing and using data. Such functional or utilitarian techniques are the subject of patent law. A survey of the United States Patent and Trademark Office (USPTO) database uncovers several hundred patents pertaining to the processing of data and the manipulation and use of databases. Patents owned by companies like Google, Microsoft, Yahoo, and several smaller Internet-based and computer outfits cover searches of databases, extraction of data, coordination and linking of databases, and other ways in which data can be arranged, coordinated, selected, and transformed. Professor Pamela Samuelson has argued that patents on data interfaces impede the interoperability of data and information to transfer mechanisms across systems, applications, and platforms. The following discussion explains how patent law is used to define and protect property rights in the data economy.

A patent protects a novel, useful, and nonobvious invention. An invention is defined in the United States Patent Act as “a process, machine, manufacture, or

---

129 Id. at 154; see Feist Publ’ns, Inc. v. Rural Tel. Serv. Co., 499 U.S. 340, 361 (1991) (stating that to prove any infringement you must prove ownership of a copyright).


131 A search of patents in the USPTO online database uncovered 94,375 patents that had the word “data” or “database” in the title and in the abstract. The search was most recently conducted on September 24, 2012.


134 35 U.S.C. §§ 101–103 (2006). A patent is obtained after an application and review process with the USPTO. Id. § 131. The administrative process checks to see if the invention is novel, nonobvious, useful, and a patentable invention. Id. §§ 101–103.
composition of matter.\textsuperscript{135} Case law establishes three exclusions from the definition of invention: abstract ideas, laws of nature, and natural phenomena.\textsuperscript{136} Data itself would not be patentable since it does not fall into any of the four categories of invention. Processed data in the form of information or a database would also not fall under patentable subject matter.\textsuperscript{137}

Tools to extract, aggregate, manipulate, or process data would, however, be considered patentable inventions. A general-purpose machine, such as a computer, can be patented.\textsuperscript{138} Similarly, a special-purpose machine, such as a heart monitor that collects, extracts, or analyzes special types of data, such as health indicators, can be patented.\textsuperscript{139} Furthermore, processes for manipulating or transforming data that are not necessarily tied to or operationalized through a machine can be patented. So methods for searching or extrapolating data can be patented.\textsuperscript{140}

A critical question with any type of patent is that of validity. The USPTO grant of a patent creates a presumption of validity, but this presumption can be rebutted in the context of patent litigation.\textsuperscript{141} A patent challenger must show that the invention is not subject matter covered by the statute or that the invention is not novel, nonobvious, useful, or disclosed in the patent specifications. For patents pertaining to data, such as the ones discussed above, the critical concerns for patent validity would be whether a particular invention is novel and nonobvious and whether a particular invention would be a process, machine, manufacture, or composition of matter.\textsuperscript{142}

For novelty or nonobviousness, the validity question would rest on the prior art in the relevant field within which the data-related invention falls.\textsuperscript{143} The novelty analysis would focus on whether an exactly identical invention already existed in

---

\textsuperscript{135} Id. § 101.
\textsuperscript{137} See Baker v. Selden, 101 U.S. 99, 102–03 (1879) (distinguishing between expressive information that is the subject of copyright but not patent and nonexpressive processes and system that are the subject of patent but not copyright).
\textsuperscript{138} See In re Alappat, 33 F.3d 1526, 1545 (Fed. Cir. 2003).
\textsuperscript{139} See id.
\textsuperscript{140} See id.; State St. Bank & Trust Co. v. Signature Fin. Grp., Inc., 149 F.3d 1368, 1373 (Fed. Cir. 1998).
\textsuperscript{142} See, e.g., In re Bilski, 545 F.3d 943, 976–77 (2d Cir. 2008) (Newman, J., dissenting) (identifying novelty, nonobviousness, and patentable subject matter as issues in the patenting of processes).
\textsuperscript{143} 35 U.S.C. §§ 102–103. For the most part, the prior art for data-related inventions will be from the fields of computer science or electrical engineering. See, e.g., In re Lowry, 32 F.3d 1579 (Fed. Cir. 1994) (involving patent for invention of a data processing system). Each of these fields will establish the state of knowledge on electrical processes, computer flow charting, and data processing that will aid in assessing the invention.
the prior art. The nonobviousness analysis would focus on whether the differences between the patented invention and what was known in the prior art would be obvious to a person having ordinary skill in the art. In other words, do the differences constitute trivial variations or substantive progress in the field? In the case of nonobviousness, data-related patents may be subject to vulnerability as there may be very few new methods of processing data that are truly innovative. However, the specific inquiry will depend on the field and the scope of the prior art.

The Supreme Court’s recent decision in *Bilski v. Kappos* has direct relevance to data-related patents. At issue in *Bilski* was what types of processes would constitute a patentable invention. The USPTO rejected a patent application disclosing a method for hedging financial risk. The disclosure did not limit the method to a particular machine or physical context. The patent examiner, following the approach of the European Patent Office, rejected the application on the grounds that there was no technical effect disclosed in the invention. In other words, the disclosure merely described a series of steps but failed to disclose any operational or engineering steps. The Board of Patent Appeals and Interferences, the first level of review within the USPTO, affirmed the rejection, but it based the rejection on the grounds that the invention was an abstract idea, as opposed to a concrete application.

On appeal, the United States Court of Appeals for the Federal Circuit upheld the rejection but on substantially different grounds. Reviewing the Supreme Court precedent on the patentability of processes—particularly software-related processes—the Federal Circuit concluded that not all processes are patentable. The court held that in order for a process to be patentable it either had to be machine-based or constitute a transformation. The difficult part of this “machine or transformation” test is discerning when a process constitutes a transformation. Partly, a transformation involves something physical, such as a change from one

---

144 See 35 U.S.C. § 102. The novelty inquiry rests on a fairly broad understanding of the prior art; it can be from any field.


146 130 S. Ct. 3218 (2010).

147 Id. at 3221–23.

148 Id. at 3224.

149 Id.

150 Id.

151 Id.

152 In re Bilski, 545 F.3d 943, 952–53 (Fed. Cir. 2008).

153 Id. at 956.

154 Id. at 957.
physical state to another as in a chemical or biological process. But a transformation could also involve a change that is nonphysical, such as the transformation of data from one form to another through mathematical or other operations. By applying the risk-hedging method to the patent at issue in Bilski, the Federal Circuit ruled that the invention did not constitute a patentable process because the hedging method at issue did not involve a machine and did not involve a transformation of data from one state to another. The process disclosed was purely mental or cognitive.

Bilski appealed, and the Supreme Court granted certiorari. Upon its review of the case, the Court affirmed the rejection and adopted a different line of reasoning. The Court held that processes in general were patentable unless they constituted an abstract idea, a law of nature, or natural phenomena. The machine or transformation test was not the sole test to determine whether a process would qualify as an invention; instead, it was one of many possible tests. Finally, the Supreme Court concluded that the hedging method was an abstract idea and therefore excluded from the meaning of an invention.

In light of Bilski, current data-related patents may face challenges because of the new standard. The Bilski opinion is opaque and reveals certain fault lines in the Roberts Court. The Court’s opinion employs a broad reading of “process,” implying that most data-related patents would be found valid. However, four of the nine justices (Stevens, Sotomayor, Breyer, and Ginsburg) would have categorically excluded business-method patents from patentability. Since many data-related patents are types of business-method patents, a broad exclusion for business methods would make data-related patents vulnerable. The Supreme Court, however, would not likely recognize such a broad exclusion, especially after Justice Stevens’s retirement.

A more realistic threat to data-related patents would be the interpretation of Bilski by the USPTO and the Federal Circuit. Subsequent to the Bilski decision, the

---

155 Id.
156 See id. at 958 (“[E]ven though a fundamental principle itself is not patent-eligible, processes incorporating a fundamental principle may be patent-eligible.”).
157 Id. at 963–66.
158 Id.; see also In re Comiskey, 554 F.3d 967, 981 (Fed. Cir. 2009) (rejecting the arbitration method patent as invalid subject matter because the method was an abstract idea and mental step).
160 Id. at 3225.
161 Id. at 3227.
162 Id. at 3231.
163 Id.
164 Id. at 3249 (Stevens, J., concurring) (arguing that business methods are not patentable subject matter).
165 Id. at 3228–29 (majority opinion) (emphasizing the Court’s reluctance to embrace categorical exclusions).
USPTO issued proposed rules for analyzing patentable subject matter.\textsuperscript{166} Future developments will show how far these limitations will be taken. What is established, however, is that patents are an important type of intellectual property for the protection of methods of data processing, which complement the copyright protection of information and databases.

\textit{(c) Trade Secret}

Information, data, databases, and related inventions can be protected as trade secret. Complementing both patent and copyright law, trade secret law provides a comprehensive form of intellectual property protection, grounded mostly at the state level and through criminal enforcement by the Department of Justice at the federal level.\textsuperscript{167} Given its breadth, trade secret may serve as a substitute for patent and copyright. Despite the breadth of trade secret law, intellectual property owners may prefer patent and copyright to trade secret protections for many reasons. These reasons include access to federal courts and the absence of strong defenses to infringement such as reverse engineering, which is a critical feature of trade secret law.\textsuperscript{168} Stated succinctly, trade secret offers broader coverage of potentially perpetual duration; yet, it is subject to stronger defenses for alleged infringers, compared to patent and copyright.

Trade secret protects any type of information whose value comes from the fact that it is not generally known and easily discernible and that a competitor does not have access to it.\textsuperscript{169} The word “information” is used in most trade secret statutes and should not be confused with the usage in this Article. The concept of information in trade secret law is broad-ranging and includes processes, formulas, and data.\textsuperscript{170} For example, customer lists are protected as a trade secret in all

\begin{itemize}
\item \textsuperscript{169} UNIF. TRADE SECRET ACT § 1(4) (1985) (defining a trade secret as “information, including a formula, pattern, compilation, program, device, method, technique, or process, that: (i) derives independent economic value, actual or potential, from not being generally known to, and not being readily ascertainable by proper means by, other persons who can obtain economic value from its disclosure or use, and (ii) is the subject of efforts that are reasonable under the circumstances to maintain its secrecy”).
\item \textsuperscript{170} See, e.g., Ruckleshaus v. Monsanto, Co., 467 U.S. 986, 1003–04 (1984) (finding that environmental data submitted to government could be protected as a trade secret).
\end{itemize}
jurisdictions, individual pieces of data, such as a price or the name of a customer, can be protected as a trade secret. Therefore, unlike with patent and copyright, trade secret can protect data itself and not just simply functional methods for processing data or expressions interpreting or compiling data.

Although trade secret covers a wide subject matter, including data, there are several limitations on the role of trade secret in commercializing data. First of all, application of trade secret depends on the secret’s value in a commercial or market context. Although anyone who breaches a trade secret, whether a direct competitor or not, can be liable for trade secret misappropriation, the protection of trade secret law extends only to commercially valuable secrets. Issues of privacy and security, which are also sources of value, are not relevant to trade secret protection. Therefore, trade secret law protects commercial interests in data, as opposed to privacy interests.

Second, trade secret protection does not extend to readily ascertainable information. Data like birthdates, names, addresses, and market variables may be excluded from trade secret protection if they are each ascertainable from independent sources. For example, owners and operators of social networking sites could not protect such readily ascertainable information through trade secret law if the information is available from independent sources. Furthermore, if such information were contained in public and readily searchable posts, then the data content would be excluded from trade secret protection.

Third, trade secret protection requires the purported trade secret owner to take affirmative and reasonable steps to maintain the secrecy of the protected information. Such reasonable steps could include technological measures—such as encryption or password protection—or contractual measures—such as nondisclosure agreements or covenants not to compete. Reasonable steps could also include the creation of special relationships that impose a duty of confidentiality on the parties. This requirement poses particular difficulties for

---

171. See, e.g., Pub. Sys., Inc. v. Towry, 587 So. 2d 969, 973 (Ala. 1991) (“It is undisputed, as a general proposition of law, that customer lists may, in proper circumstances, be afforded the protection of a trade secret.”).


173. See UNIF. TRADE SECRET ACT § 1 cmt.


175. UNIF. TRADE SECRET ACT § 1 cmt.

176. Id.


the use of trade secret law in the commercialization of data. Often, other users create data independent of any relationship with the trade secret owner. Therefore, it would be impossible for the trade secret owner to maintain the secrecy of the generated data. Consider social network sites where anyone can join, and posts can be readily searched and copied. A term of use for the site that imposes confidentiality requirements on the data and information generated on the site would be difficult to enforce and would face the same technical and social challenges as protection for personal privacy. While reasonable steps are a loose, perhaps even formal, requirement, the vulnerability of data generated by multiple users would be grounds for questioning whether this requirement for trade secret protection has been met.

\[(d)\] **Misappropriation**

Despite its uncertain roots, the common law tort of misappropriation provides a quasi-property right for certain types of data. Traced to the Supreme Court’s decision in *INS v. Associated Press*,\(^{179}\) this tort allows the aggregator, collector, or disseminator of certain types of data to prevent a competitor from improperly accessing and using the data for a limited period of time.\(^{180}\) The impropriety is based on the actions of the party who has allegedly performed the misappropriation and focuses on the extent to which that party has attempted to free ride off the efforts of the claimant.\(^{181}\) Courts typically focus on whether a party has simply copied someone else’s data or has added value in how the data is generated or used. In this way, misappropriation extends protection to items that would not ordinarily be protected by trade secret law.\(^{182}\)

Timing is key in a misappropriation action. The data protection must fall into the category of “hot news”—data that is time sensitive and whose commercial value expires quickly because it is readily leaked to the public or readily discoverable by independent sources. In the *Associated Press* case, protection extended to news reports that were gathered from the battlefields of Europe.\(^{183}\) In other cases, hot news included sports scores or market prices.\(^{184}\) Unlike trade secret

\(^{179}\) 248 U.S. 215 (1918).

\(^{180}\) *See* Nat’l Basketball Ass’n v. Motorola, 105 F.3d 841, 845 (2d Cir. 1997) (presenting a multifactor test as basis for valid misappropriation claim, where factors included timeliness of information, effort expended to obtain information, free riding in copying information, and competition between plaintiff and defendant).

\(^{181}\) *Id.* (discussing requirement of free riding).


\(^{183}\) *Associated Press*, 248 U.S. at 231.

\(^{184}\) *Nat’l Basketball Ass’n*, 105 F.3d at 854. For a successful misappropriation claim involving stock indices, see *Board of Trade v. Dow Jones & Co.*, 456 N.E.2d 84 (Ill. 1983).
law, misappropriation is designed to protect the labor of the creator in collecting, aggregating, and manipulating data.

What distinguishes the tort of misappropriation from the other types of intellectual property described in this section is its scope. The tort protects specific, time-sensitive data that may not qualify as a trade secret, a patentable invention, or copyrightable expression. Furthermore, the theory of the tort is the protection of labor invested into the generation or manipulation of the data, much as with the European Database Directive.\(^{185}\) Finally, the tort creates a quasi-property right since it only provides protection against misappropriation by a direct competitor. As applied to the commercialization of data, the tort provides highly specialized protection that might fill a gap left open by copyright, patent, and trade secret.

**(e) The Normative Foundations of Data Commercialization**

The ownership structure of data reveals an overlap of the three normative perspectives described above. The emphasis on property protection to promote creation as well as management of data is consistent with the liberal perspective. Allowing some property protection for information, processes, and secrets promotes participation in both the market and political spheres. The construction of data and information as property makes no distinction between the political and economic uses of information. Furthermore, the relevant open access given to data under the various regimes reflects the existence of a public sphere that is exempt from both economic commercialization and private secrecy. Access to data reflects a form of transparency fundamental to both politics and markets. In fact, the only area where secrecy of data is countenanced is when such secrecy permits individual economic gain, a result consistent with the liberal perspective on economics and politics.

The autonomy perspective lends support to the ownership structure in a different light. The relatively free accessibility of data reflects a domain where individuals are free to operate without the intrusion of commerce or politics. The commercial sphere encloses a domain of data, which can be understood as refined and processed data, thereby permitting commercial exploitation. Concerns of privacy and noncommercial space seem largely absent from the ownership structure, but may be implicit both in the scope of the public domain and in the distinctions among the different ways in which data are protected. Copyright and patent, and to a certain extent trade secret and misappropriation, protect highly refined data and the tools for refining data. The result is a middle ground between truly raw data and highly refined data where personal space operates through protection of individual autonomy.\(^{186}\) One potential conflict between the liberal and autonomy perspectives is that this middle ground simply may not exist for the

---

\(^{185}\) See supra text accompanying note 119 and discussion infra at Section IV.A.3.

\(^{186}\) This middle ground will be explored more in the section below on sharing and selling data. See discussion infra Section III.B.
liberal perspective, as all data is divided between that which is raw and open, and that which is refined and proprietary.

Finally, the fairness perspective casts this ownership structure in terms of competing sources of power, both political and economic. Allowing data to be openly accessible creates a zone free from the exercise of commercial and political power. Data transparency serves to ensure equal access and a commons free from proprietary influence. A proprietary zone is appropriate to allow individuals to process and hence commercialize data, but this zone should be narrowly construed. The fairness perspective would deviate from the liberal and autonomy perspectives on the need for law to protect the distribution of data and information among market and political agents. On this point, the fairness perspective would ask more of the ownership structure than has been described at this point. It would call for more internal limiting doctrines (such as fair use and the first sale doctrine) as well as external limiting doctrines (such as laws protecting privacy, unfair competition, and misuse of personal information). In short, the fairness perspective would ground greater data transparency as well as limits on data commercialization.

An important constitutional background to the commercialization of data is the protection given to freedom of speech under the First Amendment of the U.S. Constitution and comparable protections in other jurisdictions. The First Amendment is in the background, interjecting in only special cases. From the liberal perspective, the ownership structure complements the First Amendment by creating a market for expressions that complements the marketplace of ideas. Economic freedom and political freedom go hand in hand to protect full citizen participation in the economic and political realms. In contrast with the liberal perspective, the autonomy perspective considers the First Amendment to be limited by concerns for privacy and security in the middle ground that arises between raw data and highly refined data. Finally, the fairness perspective uses the First Amendment as a limiting principle to protect against power imbalances in

---

187 See Pamela Samuelson, The U.S. Digital Agenda at WIPO, 37 VA. J. INT’L L. 369, 381 n.74 (1997) (discussing fair use and first sale doctrines as important limits on ownership rights over information, thereby protecting access).


189 See, e.g., IMS Health Inc. v. Sorrell, 630 F.3d 263 (2d Cir. 2010) aff’d, 131 S. Ct. 2653 (2011) (involving First Amendment challenge to state statute prohibiting the sale, license or exchange of pharmaceutical data).

190 As examples of this view of the First Amendment, see Eldred v. Ashcroft, 537 U.S. 186 (2003) (finding narrow First Amendment limits on copyright law), and Sorrell v. IMS Health, Inc., 131 S. Ct. 2653 (2011) (affirming First Amendment claim to allow for commercialization of data).

191 See, e.g., IMS Health Inc. v. Ayotte, 550 F.3d 42 (1st Cir. 2008) (upholding statute limiting commercialization of pharmaceutical data against First Amendment challenge).
the various spheres in which data exists. 192 Speech serves to protect politically or economically disadvantaged groups from the misinterpretation and misuse of data and the generation of information that can be misleading and therefore politically and economically harmful. 193

In short, the ownership structure over data is both descriptively and normatively complex. Each of the normative perspectives plays a role in understanding the governance of data. A similar complexity can be seen in the government ownership of data and in the notion of open records.

2. Government Data and Open Records Acts

Governmental bodies, including agencies, generate data and information. For the most part, the rules of intellectual property law apply to such data and information—although the particular context of government ownership shapes the specific application of the intellectual property rules. For example, federal government bodies cannot own copyright in materials under the United States Copyright Act unless an independent contractor creates the materials and transfers copyright to the government. 194 There are no statutory restrictions on ownership of copyright by state governments, although some jurisdictions have created restrictions for state legislative and regulatory materials. 195 As far as patent law goes, governments have not obtained patents in data- or information-related inventions, although they could. Governments do, however, own trade secrets in data and information, which has been the source of controversy for proponents of open government. 196 Finally, misappropriation would exist as a cause of action for governmental entities, although no reported cases illustrate such claims.


193 See, e.g., Houchins, 438 U.S. at 12 (discussing interests in protecting inmate population from news reporting and gathering).

194 For an analysis of exclusions from intellectual property protection of legal materials, see Shubha Ghosh, Copyright as Privatization: The Case of Model Codes, 78 TUL. L. REV. 653, 663–68 (2004) (analyzing importance of excluding statutes, regulations, and case law from copyright protection from a democratic governance perspective).

195 See, e.g., Veeck v. S. Bldg. Code Cong. Int’l Inc., 241 F.3d 398 (5th Cir. 2001) (holding that privately created model codes retained their copyright protection, even after adoption by state agencies).

Government-owned data and databases pose challenges for public accountability and access. All fifty states have enacted open records acts, which make governmental entities the custodian of records with the obligation of disclosing certain records upon a proper request.  

Open records acts are often considered the product of Progressive Era legal reform, designed to make government more accountable and accessible to the public. However, many states adopted open records acts prior to the period known as the Progressive Era—roughly the period from 1890 to 1920. The State of Wisconsin, for example, often associated with quintessential Progressive Era reforms, enacted its Open Records Act in 1850, shortly after its inclusion into the United States. The first version of the Act was an absolute open statute, meaning that most of the restrictions on access were procedural ones. Over time, however, courts placed reasonable limits on access to protect interests such as privacy, property, or security.  

An analysis of the current version of Wisconsin’s Open Records Act highlights many of the themes of this Article and raises additional questions. The statement of purpose enunciated in the Act illustrates both its bold spirit and some potential tensions:  

[I]t is declared to be the public policy of this state that all persons are entitled to the greatest possible information regarding the affairs of government and the official acts of those officers and employers who represent them. Further, providing persons with such information is declared to be an essential function of a representative government . . . .  

The provision concludes by creating a “presumption of complete public access” with access being denied “only in an exceptional case.”

Terminology is critical to the application of the Open Records Act. The purpose statement pertains to access to “information,” a term that is not defined in the statute. The word “data” is not used in the statement of purpose, although it does arise in another section discussed below. Presumably, the word “information” has its ordinary meaning and not the specialized meaning examined in this Article. This ordinary meaning would include not just interpretations of raw data, but facts arising in governmental proceedings. The key concept, however, is neither information nor data, but a record. Needless to say, the purpose of the Open Records Act is not simply access to information, but recorded information. The definition of “record” is pivotal to the application of the Act.

There is no obligation to create a record under the Open Records Act. But if a record is kept, then it is subject to accessibility. A record is defined as “any material on which written, drawn, printed, spoken, visual or electromagnetic information is recorded or preserved, regardless of physical form or characteristics, which has been created or is being kept by an authority.” The authority referred to in the definition is a governmental authority. As we will see in one case study discussed in Section IV, this authority can also include a nongovernmental authority to which power has been delegated by a governmental authority. The definition speaks broadly in terms of the medium and the potential types of information that would constitute a record. But the definition excludes copyrighted and patented materials and published materials that are on sale or available in a public library. Also excluded are drafts, notes, and other preliminary materials used to create a record. These exclusions protect certain property and economic interests as well as the ability of an authority to create a record without having to disclose the thought process leading up to the record’s creation. A record is a statutorily recognized and official version of the information (and presumably data) that the Open Records Act makes accessible to a requesting public.

Exclusions also exist for requests for records under the Act. Specifically, the statute excludes law enforcement records, contractors’ records, trade secrets, computer data, and certain other types of personal and proprietary data. This exclusion applies to the request, but not to the definition of “record,” which means that the authority subject to a request can deny access if the request is for what is

---

204 Id.
205 See WIS. STAT. ANN. § 19.32; infra text accompanying note 207 (defining “record” in Wisconsin’s Open Records Act).
206 See Kidwell, supra note 198 at 1026–28 (drawing a distinction between a record and copyrighted material).
207 WIS. STAT. ANN. § 19.32(2).
208 Infra Part IV.A.1.
209 WIS. STAT. ANN. § 19.32(2).
210 Id.
211 Id. § 19.36.
deemed to be a trade secret or restricted data. This distinction is important because the inclusion of these items as a record means that the authority must serve as a custodian and maintain these items. There is, however, authority to exclude these items from a record request.

Government ownership of data has three different dimensions, each reflecting the normative complexity of the legal structure of data ownership: (1) the role of the government as a market actor, (2) archiving data and information, and (3) transparency and openness. Each of these three dimensions reflects the normative complexity in its own way.

Government control of data reflects the government’s role as a market actor, both through intervention in market activity and through participation. As an intervenor, the government obtains data from regulated entities but also manipulates and interprets such data through the regulatory process. When the government is an active market participant, through state-owned enterprises, the government also participates in the generation and accumulation of data. Such a role is consistent with the liberal perspective that supports strong overlap between the market and political realms. The autonomy perspective sees conflict, however, especially if the government’s role as regulator feeds into its role as market participant—perhaps through state-owned enterprises’ use of data obtained through the regulatory process. Finally, the fairness perspective is concerned with the undue influence of the state in the private realm through its coercive and potentially monopoloid power. As the case studies in Section IV will show, these distinct normative perspectives can overlap and conflict in practice.

The archiving dimensions of open records acts also reflect this normative tension. The liberal perspective considers the archiving function to be part of the dual market and political realms of liberal society. Open records provide a source of information that can aid in political deliberation and market participation. Arguably, the website Data.gov exemplifies the liberal perspective where data accessibility provides a way to gain knowledge that is useful in the political realm with minimal limits on commercialization. But both the autonomy perspective

---

212 See infra Part IV.A.2. Open records acts provide an important comparison with Data.gov. Although both ostensibly serve the goals of public access, each illustrates different approaches. Open records acts serve as a means to impose obligations on government to maintain and allow access to recordings that are made of official proceedings. The acts serve an archival purpose that documents the actions of governmental entities. A website like Data.gov also serves an archival purpose of sorts, but its design is to allow individuals to access and use data that is collected or held by the government. Neither requires nor mandates the creation of a record or the accumulation of data. The presumption is that these activities will occur anyway. The two mechanisms are created to allow access of sorts to the records and data created. The open records acts can be compared to a museum or a library whose collection can be accessed under certain circumstances. A website like Data.gov, by comparison, is like the Exploratorium, which allows attendees to actually touch, play with, and otherwise use the items collected. Section IV will explore these differences in greater detail in the context of case studies.
and fairness perspective support restrictions on data archiving. Individual autonomy involves remembering as well as forgetting the past, and the archiving of data provides a trail of one’s past that one might prefer to expunge. Similarly, the fairness perspective draws attention to power imbalances regarding how data is recorded and stored. The rich can have records expunged or controlled while the poor may lack this capability.

Finally, the normative tensions are apparent in the transparency and openness provided by both open records acts and websites like Data.gov. The liberal perspective heralds transparency, echoing the sentiments of Justice Brandeis that sunlight is the best disinfectant.213 Once again, the autonomy and fairness perspectives invite caution: to whom, and for what purpose, do open records laws provide transparency? The autonomy perspective is concerned with who is allowed access to records and for what purposes. For example, those adhering to the autonomy perspective might find it intrusive to provide journalists or paparazzi access to records revealing personally embarrassing facts. The fairness perspective is concerned with equality issues that are overlooked in open records acts, which emphasize values of freedom over equal distribution of resources for access to and use of records that may contain sensitive data and information.

Government data ownership is a critical part of the legal ownership of data. These issues will be explored in greater depth through the case studies in Section IV, particularly those involving access to court records and real estate assessments on Data.gov. In addition to issues arising from the ownership of data, an understanding of the legal structure governing data must also account for issues arising from data transactions.

### B. Selling, Sharing, and Commercializing Data

Legal ownership of data is necessary to perform market transactions of data; meanwhile, other legal rules may further limit those transactions. Such legal rules are designed to protect the interests of transacting parties. The legal structure of transactions tracks with the rules of contract, intellectual property licensing, tort law, and other regulatory fields. And there are several specific structures worth highlighting here.

First, U.S. law allows for restrictions on the use of data through contract law. In ProCD, Inc. v. Zeidenberg,214 the Seventh Circuit held that a seller of a database could restrict access to data through a contractual limitation.215 Such a contract term was found not to be preempted by federal intellectual property law and was

---

213 LOUIS D. BRANDEIS, OTHER PEOPLE’S MONEY: AND HOW THE BANKERS USE IT 92 (1914) (“Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.”);
214 86 F.3d 1447 (7th Cir. 1996).
215 Id. at 1449, 1455.
held to complement the interests of intellectual property holders.\textsuperscript{216} This decision has been a controversial one since it was first handed down in 1996, but with some narrow exceptions regarding contractual formation, the ruling is generally accepted.\textsuperscript{217} The rationale behind this contractual protection of data is that the restriction on data access applies only to the parties to the specific transaction and therefore does not conflict with the copyright principle that data cannot be copyrighted.\textsuperscript{218} Furthermore, the contractual provision is deemed to protect the expectations of the party, allowing the seller to turn the data into a marketable commodity from which economic value can be extracted beyond the property right in the database.

Whether such a rationale is consistent with intellectual property policy has been a central controversy. Even if the restriction is between the transacting parties, mass market licensing effectively binds any purchaser to the contractual term restricting data. Therefore, what seems ostensibly to be a contract right becomes a property right, inconsistent with copyright law. Admittedly, the remedies may be different for a contract breach than for copyright infringement, but the practical reality is that purchasers will not feel the difference between being bound by contract and being bound by copyright law. In addition, the use of arbitration—which is not as constrained by formal legal rules and precedent as adjudication—to enforce the term may limit the difference in remedies and thereby further diminish the difference. An open and critical question, which will be taken up in Section IV, is the extent to which technological protections, such as encryption, can be used to protect data.

In addition, legal rules attempt to protect against privacy invasion and interference with personal interests. As many scholars have noted, such protections are piecemeal, often dependent on the type of data. The Health Insurance Portability and Accountability Act of 1996\textsuperscript{219} (HIPAA) regulations provide one example of detailed government regulations protecting privacy interests in health data.\textsuperscript{220} The tort of intrusion generally protects against invasions of one’s reasonable expectation of privacy,\textsuperscript{221} and the related tort of false light prevents reputation-harming uses of data.\textsuperscript{222} The right of publicity, which is one of the original common law torts of privacy, has taken on a life of its own and protects

\textsuperscript{216} Id. at 1453–55.
\textsuperscript{218} See ProCD, 86 F.3d at 1454–55.
\textsuperscript{220} See SOLOVE, supra note 8, at 69–70 (analyzing HIPAA).
\textsuperscript{221} RESTATEMENT (SECOND) OF TORTS § 652B (1977).
\textsuperscript{222} Id. § 652E.
against the unauthorized commercial exploitation of one’s public persona.223 For all of these torts, the First Amendment places some limitations on tort claims against reporting or communication of private data and information.224 Finally, the Fourth Amendment would place limits on the government, whether acting in its crime enforcement or its regulatory capacity, regarding unwarranted searches and seizures and data and information obtained directly or indirectly from such searches or seizures.225

With advances in genetic technologies, the potential for the misuse of personal genetic information has grabbed the attention of scholars and policymakers. Statutes prohibiting genetic discrimination prevent unequal treatment in the use of genetic data.226 Collection of tissue samples and cell cultures are not governed by property law,227 but would be regulated by tort law. The famous case of John Moore, whose tissue line was extracted by medical researchers at UCLA Medical School, brought harvesting of personal genetic information under legal scrutiny.228 Although one cannot claim a property right per se in one’s body or genetic materials, one can bring claims of lack of informed consent and breach of fiduciary duty against researchers who misuse such materials.229 Such potential claims shape transactions in genetic data and the increased use of informed consent in hospital and medical research settings.230 The case of Henrietta Lacks received scholarly and media attention.231 A cell line obtained during Lack’s treatment for ovarian cancer had been used without her consent by several generations of medical researchers and students after her death.232 Her misfortune illustrates broader trends in personalized medicine and the critical role of personalized medical and genetic information in the context of the patenting and marketing of medical diagnosis and treatment methods.233

223 See id. §§ 652C, 652D.
227 See Wash. Univ. v. Catalona, 490 F.3d 667, 673–74, 676 (8th Cir. 2007) (treating biological materials as inter vivos gifts from research participant to university).
228 Moore v. Regents of the Univ. of Cal., 793 P.2d 479 (Cal. 1990).
229 See, e.g., id. at 483, 497.
230 See SKLOOT, supra note 16, at 202–03 (describing benefit-sharing agreement between medical researchers and patient permitting harvesting of patient’s biological materials).
231 See id. at 218–19.
232 See id. at 215–16.
233 See Ghosh, supra note 16, at 475.
The legal structure of data-based transactions illustrates the same normative complexity as the legal structure of data ownership. The three normative perspectives surface and justify different parts of the legal structure. The liberal perspective supports a legal regime that protects data through contractual restrictions. The autonomy and fairness perspectives provide support for several limitations on data-based transactions, especially through protections for privacy and against the exploitation of genetic data and information.

IV. COMMERCIAL SPEECH AND COMMERCIALIZING INFORMATION

This section brings together the economic, legal, and normative issues raised by the commercialization of data through three case studies and a reconsideration of the Sorrell decision. Each case study demonstrates how goals of transparency and openness come into tension with the pursuit of commercialization. The case studies show how the issues that arise from commercializing data go beyond the facts of Sorrell. After an analysis of the case studies, the discussion returns to the Sorrell decision to explain how the richer understanding of the marketplace for information enhances the narrow conceptualization of the marketplace of ideas. The section concludes with a proposal for commercializing data that addresses the normative concerns discussed in this Article and that passes muster as regulation consistent with the view of the First Amendment underlying the Sorrell decision.

A. Three Case Studies: Real Estate Data, Government Data, and European Databases

This subsection presents three case studies. The first has to do with access to property-assessment data collected by municipalities through private corporations. This case study focuses on the transparency of data collected by the government and the use of such data for commercial purposes. The tension between governmental transparency and commercialization illustrates how states and markets—public and private institutions—overlap in the creation and management of data. The second case study examines the interaction between public and private in terms of access to federal government data and transparency at the federal level. This case study shows how regulation controls the access and use of federal data. Finally, the third case study examines the treatment of data and databases under the European database initiative and shows how a regime designed for the proprietary treatment of data and databases has shifted to one that limits private ownership. Together the three case studies illustrate the economic, legal, and normative issues framed in the previous sections. Furthermore, these case studies provide a foundation for the licensing model to be presented in the final subsection.

234 See discussion infra Part IV.A.1.
1. Real Estate Data

In 2001, WIREdata, a private company providing information to realtors on real property that is for sale, submitted a request under Wisconsin’s Open Records Act to various municipalities for tax and real property assessment data. The goal was to include the data in online real estate listings so that potential buyers would have more information about properties that might be of interest for purchase. Some municipalities informed WIREdata that they did not have the data, which had been collected by Assessment Technologies, a private company that compiled and archived the assessment data. WIREdata requested the data from Assessment Technologies, but the company denied the request on a number of grounds, including the rationale that the Open Records Act did not apply to private entities and that even if the Open Records Act did apply, the assessment records were exempt from the Act because they were protected by copyright law. WIREdata filed a mandamus action in Waukesha County Circuit Court against Sussex and Assessment Technologies, and Assessment Technologies filed for an injunction in the U.S. District Court for the Eastern District of Wisconsin to prohibit WIREdata from infringing its copyrights. The U.S. District Court ruled in favor of Assessment Technologies, holding that Assessment owned the relevant copyright in the assessment software and any derivative works. WIREdata appealed successfully to a panel of the Seventh Circuit, which ruled in a unanimous opinion by Judge Richard Posner that copyright law protected original databases but not the data of which the database is composed. The Seventh Circuit held that using copyright law to deny access to unprotected data constituted copyright misuse, an equitable doctrine that limits the right of a copyright owner who attempts to extend the scope of copyright. The effect of the appellate ruling was that Assessment Technologies could not refuse WIREdata’s request on the grounds that the records were protected by copyright.

The Seventh Circuit’s ruling set the stage for the final acts in the drama of WIREdata’s search to obtain what it viewed as the holy grail: real estate data that could be electronically manipulated. WIREdata had requested electronic copies of the assessment data so that they could be readily entered into its real estate information website. In light of the Seventh Circuit’s ruling, Assessment Technologies decided to comply with the request for electronic copies of the data.

---

235 See WIREdata, Inc. v. Vill. of Sussex, 751 N.W.2d 736, 741 (Wis. 2008).
236 Id. at 742, 744.
237 See id.
238 Id. at 744.
239 Id. at 744–45.
241 Id. at 646–47.
242 See WIREdata, 751 N.W.2d at 743.
by providing electronic files of the assessments in PDF format. While this format was technically electronic, the PDF files were not ones that anyone could readily manipulate to include on the web and link to existing real estate files. WIREdata brought suit against Assessment Technologies and several municipalities in Wisconsin for failure to comply with the Open Records Act.

In 2008, the Wisconsin Supreme Court gave WIREdata a mixed victory. The high court held that the municipalities did not deny a request for records by referring the requestor to a private third party, and the mandamus action was therefore inappropriate. But because the third party was not an authority under the Act, they were not subject to the open records law; therefore, the municipalities remained liable for any damages or fees resulting from noncompliance. However, the Wisconsin Supreme Court also ruled that the transfer of the PDF files satisfied WIREdata’s request for electronic records since the files were in the requested electronic form. In other words, the high court held that the municipalities—not Assessment Technologies—were required to provide the data sought in the original request and that the municipalities had fulfilled that requirement.

Among the issues raised in WIREdata is the mixed public-private partnership in the collection and dissemination of data. The municipalities, as part of their power to tax, also have the power to collect economic information that has value outside the government context. Governments outsource the information-collection function to private entities that can make use of technology and business management to more effectively create databases. The problem is one of ownership and rights in the databases that are created. Open records acts serve to create a transparent government that allows citizens to view state-created records. The state cannot escape obligations of its open records act through privatization. But these acts also impose obligations on private citizens in framing requests for records and, under some circumstances, in bearing the burden and cost of disclosing the requested information. Through this mechanism, a market-like institution is created for the generation and dissemination of data.

243 Id. at 745.
244 See id.
245 See id. at 739.
246 Id. at 752–53.
247 Id. at 757.
248 Id. at 759–60, 762.
249 Id. at 758–59.
250 See id. at 741.
When government data collection and open records obligations are understood as market-mimicking institutions, it should not be surprising that the creation of transparent government through open records acts can also support commercial, as well as political, activities such as the ones at stake in the *WIREdata* case. The Seventh Circuit’s decision is consistent with the liberal perspective on data and information access: it supports the notion that the market sphere and the political sphere overlap. Open records acts do not impose scrutiny on the motivation for a records request. The request could arise from mere curiosity or from the pursuit of a commercial interest, as in *WIREdata*. The government serves as a data aggregator and manager to satisfy the interests of citizens who can access the records. At the same time, open records laws recognize that copyright, trade secret, or patent can protect data from access requests by private citizens. This private sphere within transparent government serves to protect those who process data into economically valuable products recognized by intellectual property law. In this way, open records acts preserve a traditional, liberal notion of markets and politics, with certain data in the public sphere accessible through a political process alongside a proprietary sphere that supports markets.

This scheme of data governance overlooks several critical issues. For example, under its terms, the data submitted to the government falls into the public domain unless the data is transformed into the subject of intellectual property. But this contribution to the public domain ignores issues of privacy and reputational interests that can be compromised by misuse of the data. These concerns are exacerbated by the ability of the data to be commercialized and readily used by the public. Within the liberal perspective, subsidiary laws, such as defamation, may protect privacy and reputational interests. But such protections may not be fully adequate under an autonomy or fairness perspective because of the defenses to defamation and other reputation-related claims provided to the press. Protection of these privacy and reputational interests would require thinking beyond individual freedom to shape the normative framework within which market transactions and politics occur. The open source proposal in Section IV.B.2 addresses these concerns.

---

253 See 76 C.J.S. Records § 115 (2012) (“Under some statutes, every person has a right to inspect public records, whether or not he or she has any special need for or interest in such records, and regardless of his or her motive, purpose or intended use of the records.” (citations omitted)).


255 See, e.g., Cox Broad. Corp. v. Cohn, 420 U.S. 469, 494–95 (1975) (allowing publication of truthful statements obtained from public documents despite intrusion of privacy).
The case study above examined the interaction between public collection and commercial uses of data. The next case study examines the creation of a public information data library that is openly accessible. This case study illustrates similar problems with the liberal perspective, but also demonstrates a potential solution through regulation.

2. Government Data

While open records acts can be traced back to the nineteenth century as a mechanism for citizens to request information from the government, the contemporary trend, made possible by the Internet, is publication of government-collected data for ready access. This trend to publish data, initiated by state and federal governments, has led private entities to create search tools to identify data, aggregate it, and analyze it. Google, for example, provides Public Data Explorer, an application designed for journalists and researchers to analyze public databases to discover relationships and correlations in government-gathered data.256 Within this environment, it is not surprising that statisticians are viewed as modern-day rock stars, with their ability to make data speak and use mathematical techniques as investigative tools.

Public databases serve as resources for commercialization and the development of socially useful applications. A notable example is the use of publicly available information on train schedules to create a smartphone application to check on arrival and departure times for connecting trains.257 Likewise, software engineers explore available data on energy usage to identify areas to target conservation efforts.258 These examples illustrate how government investment in data collection and dissemination can spur private enterprise. Government-collected data are made available to the public as resources for new applications and information-based products. Unsettled issues exist, however, about ownership of commercially successful applications and compensation for use of government data. Although these ownership and compensation issues are far from resolved, disputes have frequently been resolved in the government’s favor.259

259 See, e.g., Snepp v. United States, 444 U.S. 507 (1980); Fu, supra note 257.
The Obama Administration initiated the creation of the website Data.gov, which provides access to data collected by federal agencies that is not otherwise protected as a government secret for security reasons.\(^{260}\) The website’s terms of use illustrate the current administration’s commitment to both government transparency and protection of private interests. For example, the data policy statement that is linked from the Data.gov homepage limits data access that conflicts with protections for national security and privacy as defined by the Office of Management and Budget (OMB) guidelines.\(^{261}\) The statement also imposes requirements for citation to datasets downloaded from the site and obligations imposed on the collecting government agency for data retention and data integrity.\(^{262}\) Finally, the statement places some restrictions on “secondary uses” of data downloaded from the site:

Data accessed through Data.gov do not, and should not, include controls over its end use. However, as the data owner or authoritative source for the data, the submitting Department or Agency must retain version control of datasets accessed. Once the data have been downloaded from the agency’s site, the government cannot vouch for their quality and timeliness. Furthermore, the US Government cannot vouch for any analyses conducted with data retrieved from Data.gov.\(^{263}\)

Arguably, the requirements are not terms that can be enforced as in a contract unless a claim can be made that the terms were consented to. But even if it is not enforced as a contract, the terms constitute a policy or norms for conduct with respect to government data.

Particularly worth noting about the site is the distinction drawn between the concepts of “data” and “information.” In the Frequently Asked Questions (FAQ) section, the answer to the question “What is data?” is: “Data are values or sets of values representing a specific concept or concepts. Data become ‘information’ when analyzed and possibly combined with other data in order to extract meaning and to provide context. The meaning of data can vary according to its context.”\(^{264}\) The distinction between data and information is critical here for a number of


\(^{262}\) *Data Policy Statements*, supra note 21.

\(^{263}\) Id.

\(^{264}\) *Frequently Asked Questions (FAQ)*, supra note 22.
reasons. Under the terms of the site, public access and agency maintenance obligations apply to data, but not to information.\textsuperscript{265} Also, the distinction pertains to both what users are allowed to do and how the government must carry out its duties.\textsuperscript{266}

The term “data” refers to representations of some concept or set of concepts. This representation is quantitative, as the phrase “values or sets of values” suggests, but could arguably be qualitative, as perhaps the description of a certain condition of the environment. Information is made from data, as the text suggests, by stating, “data become ‘information.’” The text implies that information is processed data, analyzed through some method, whether statistical or qualitative. Information, as processed data, is not subject to the access rules, nor is it subject to the limitations of secondary uses. A user under these terms can download data from the site, process it, perhaps through a statistical technique or qualitative analysis, and the end product of those analyses is information to which the user can limit access as long as the downloaded data is still made available.

The distinction between data and information is a critical one for the governance of the datasets made publicly available through Data.gov. The distinction allows users some proprietary protection for information generated from the publicly available data while the data remains freely accessible. There is more than a hint of Lockean thinking here since information produced from the data is the fruit of the user’s labor.\textsuperscript{267} But there is also an interpretation consistent with the values of transparency and open government. The terms of Data.gov allow individuals to use data to form their own interpretations and evaluative opinions. These interpretations and opinions gleaned from data would be classified as information over which the user might have exclusionary rights. This information becomes the tool for individual users as they engage with others in deliberation and participation in open government.

The autonomy perspective envisions separate private and public spheres that interact with each other while maintaining independence and integrity. These separate spheres have implications for the protection of personal privacy. But they also have implications for security and government secrecy. The government also engages in the processing of data. This processing of data produces information that would not be subject to the values of open data represented in Data.gov. One example of such information is national security—a category expressly recognized within the terms of the site as exempt from open access. This exemption is included for pragmatic reasons. But the data-information distinction provides

\textsuperscript{265} See id. (“Data become ‘information’ when analyzed and possibly combined with other data in order to extract meaning, and to provide context.”); Data Policy Statements, supra note 21 (describing general obligations, including Information Quality Guidelines).

\textsuperscript{266} See Data Policy Statements, supra note 21; Frequently Asked Questions (FAQ), supra note 22.

\textsuperscript{267} See, e.g., Wendy J. Gordon, On Owning Information: Intellectual Property and the Restitutionary Impulse, 78 Va. L. Rev. 149, 167 n.64 (1992) (describing Lockean Theory as the “conviction that it is unfair for one person to take the fruits of another’s labor”).
another rationale for the exclusion. National security represents information that results from the autonomous sphere of governmental decision-making, an analogue to the sphere of privacy enjoyed by individual citizens.

The autonomy perspective helps to explain some of the distinctions drawn by open government projects like Data.gov, such as the exclusion for national security and privacy. But the autonomy perspective fails at a basic level. If the boundaries of public and private rest on the distinction between unprocessed (or raw) data and processed (or cooked) data, how do we draw the line between processed and unprocessed? The terms of use assume that data collected by the government through its various mechanisms (tax collection, census, property information) is raw and unprocessed and does not become information until someone acts on it. But the collected data may be initially processed as well, its collection based on assumptions of how to define a unit of analysis, what questions to ask, and what tools should be used for measurement. Furthermore, there is a question of how much processing is required for data to become information. Is simply arranging the data in some pattern enough? Does the data have to be aggregated, statistically analyzed, or summarized in some fashion? Copyright law handles these issues through the concept of originality to determine the boundaries of proprietary rights. The line between data and information is not so clear under the terms of the Data.gov site. Rights pertaining to information are based on the background rules—copyright, patent, trade secret, and misappropriation laws—discussed in the previous sections.

Although the case study of government data may not provide a rich set of answers in its illustration of the autonomy perspective, the model of Data.gov provides a useful example that can inform the open source solution that is discussed later in this section. The terms of use from the website can serve as a template for a type of norm, seeding through open source licensing models for the use and dissemination of data. The open source licensing models, however, may support a richer normative framework than what is evinced by Data.gov.

3. European Databases

Open government initiatives in Europe parallel many of the developments in the United States, as European governments and citizens seek to make government-collected data available through web portals. Many European governments are playing catch-up with the United States through the implementation of legislation that echoes the U.S. Freedom of Information Act.
Such initiatives demonstrate that the concerns raised in this Article about commercialization of data are global and require a comparative perspective.

Two major differences between U.S. and European approaches to data commercialization qualify some of the points raised in this Article. First, the European Union has more centralized protections for individual data privacy than the United States, which treats the issue as a matter of decentralized private law rather than centralized regulation. As a result, many of the European initiatives for open government attempt to stave off the privacy issues raised by the commercialization of data. Second, the European Union, through the 1996 European Database Directive (Directive), protects databases and data more strongly than the United States, with the result that private incentives for the commercialization of data may be stronger. These two differences in approach provide a useful case study for understanding alternative regulatory structures and their effects on data commercialization.

Professors Reichman and Samuelson, in one of the first scholarly articles on the Directive, point out that one of the purposes for implementing the Directive was the promotion of database industries in Europe. A 2005 study of the Directive, based mostly on surveys of members of the database industry, concluded that the Directive and its implementation by member states had little effect on the development of the industries, with the United States still dominating over the European Union. The continuing dominance of the United States in the database industry is particularly striking because the U.S. Congress rejected the adoption of European-style database protections several times during the 1990s. Instead, the United States relies on the regulatory environment described in Section II of this Article—a mix of weak intellectual property and common law protections. The 2005 Directive study suggests that uncertainties and ambiguities over the interpretation of the rights created under the Directive are responsible for its ineffectiveness. However, the study does not address how the United States has managed to dominate globally despite also lacking strong domestic protections for database industries.

The Directive creates a sui generis right for databases constructed through the gathering of data by the first creator. Unlike copyright law, which bases protection on the originality shown in the selection, arrangement, and coordination of the data, the sui generis right is established through substantial investment in the

---

270 See Reichman & Samuelson, supra note 123, at 95–98.
271 See id. at 87–88 (describing potential threat to public domain from overprotection of data through expansive exclusionary rights).
272 See id. at 73–74.
274 See Reichman & Samuelson supra note 123, at 96–97.
275 Id. at 84–85.
gathering and collecting of data. The right prevents the extraction of data from the database or use of the database without permission of the owner. Exceptions are created for research and educational purposes. Such protections were designed to supplement copyright protections for databases and harmonize conflicting standards for protection across members of the European Union.

The 2005 study suggests that the failure of the Directive to stimulate the database industries stems in part from the exceptions created for research and educational purposes. Although there is not much litigation on the scope of these exclusions, the authors of the study report that the existence of the exclusions may create uncertainties among users. Furthermore, several European Court of Justice (ECJ) opinions may have weakened the scope of the sui generis rights. Two important opinions impose limits on the exercise of the sui generis rights based on competition law principles, specifically the denial of access to a database that might constitute an essential facility. The two cases dealt with a guide to television programming created by government-run television and a database structure for recording pharmaceutical sales. The standard enunciated by the ECJ for the imposition on a compulsory license, compelling access to data and databases, raised questions about the scope of the sui generis right created by the Directive.

Additional uncertainty arose from the court’s decision in *British Horseracing Board Ltd. v. William Hill Organization Ltd.*, involving access to horse racing data. The court ruled that the Directive did not protect an Internet database consisting of racehorse statistics because the substantial investment in the database was not from the collection and gathering of the data, but from the creation of the data. Arguably, the decision reached further than the competition law decisions because the court placed limits on what types of databases were protected by the Directive. In the *William Hill* decision, the court held that the Directive intended to

---

276 Id. at 85.
277 Id. at 91–92.
279 Id. at 21, 23.
283 See Comm’n of the European Cmtys., supra note 273, at 23.
285 Id. at I-10468.
286 Id. at I-10478, I-10494.
protect significant investments in gathering data, not the relatively insignificant investment in construing data through some statistical or interpretative process.\textsuperscript{287} The decision was particularly confusing because it suggested that the first to construct the data could not obtain protection in the database, but an entity that gathered the constructed data—perhaps by creating a database of horseracing statistics—could obtain sui generis rights in the secondary database. The 2005 study concluded that this 2004 decision helped to create uncertainty and may have undermined the purpose of the Directive to the extent it was meant to create stable and certain property rights in databases.\textsuperscript{288}

The experience with the European Database Directive illustrates some of the problems with establishing strong property rights in data. Foundational competition law and policies limit the strength of this property right. Furthermore, there are underlying and unresolvable questions of what constitutes data. These issues also arise in the context of the United States, as \textit{WIREdata} and Data.gov illustrate. The distinction that the European court drew between constructed and unconstructed data in \textit{William Hill} parallels the discussion of the demarcation between data and information in the context of Data.gov. The difference is that the Database Directive protects constructed data, while Data.gov protects information through background rules of intellectual property and common law. If the world of data commercialization is moving toward a regime of strong property rights and market protection, then the experience with the European Database Directive may be a harbinger of very real pitfalls to a system of strong legal entitlement.

A comparison of the prevailing approaches to data and databases in Europe and the United States suggests that generally open access to data and thin protection for databases may be more supportive of a healthy and competitive database industry. The openness permitted under the United States’ regime allows for competition in the provision of services, allowing consumers to access and manipulate information. Granted, the form of competition may not be the pure textbook kind whereby firms and many consumers are driven by price signals. Rather, the types of competition seen in database markets may be more of the oligopolistic type with few firms competing on different margins for consumers.\textsuperscript{289} Nonetheless, such competition can be vibrant for industry and may be consistent with other noneconomic values such as autonomy, access to knowledge, and transparency of government and market institutions. While the first and second case studies provide some insight into different types of legal entitlement

\textsuperscript{287} \textit{Id.} at I-10464, I-10476, I-10494; \textit{see also} Council Directive 96/9, 1996 O.J. (L 77) 25 (EC) (“[T]he making of databases requires the investment of considerable human, technical and financial resources while such databases can be copied or accessed at a fraction of the cost needed to design them independently . . . .”).

\textsuperscript{288} \textit{See Comm’n of the European Cmty.}, \textit{supra} note 273, at 24, 27.

\textsuperscript{289} \textit{See} Erik Meijer & Gavin Bierman, \textit{A Co-Relational Model of Data for Large Shared-Data Banks}, ACM QUEUE: PROGRAMMING LANGUAGES (Mar. 18, 2011), http://portal.acm.org/ft_gateway.cfm?id=1961297&etype=pdf (“[T]he relational database market is a classic example of an oligopoly.”).
structures and their protection for values of freedom and autonomy, the third and final case study shows how competition informs the database industry and the market for data. The three together provide background for understanding the open-source-licensing model for data, presented in the next section.

B. Sorrell Revisited and a Proposal

1. The First Amendment and the Marketplace of Information and Data

Both the Sorrell majority and dissent operate under an antiquated notion of the marketplace of ideas. According to this antiquated notion, individuals communicate with others to persuade and as part of a system of deliberation and debate. The state’s interference in this process is suspect if it attempts to close off channels of communication and thereby reduce knowledge necessary for individual decision-making. The majority ruled that Vermont’s restrictions on the sale and use of data closed off channels of communication to physicians making prescription decisions. By preventing certain data-mining practices, the State was fettering the marketplace of ideas and inhibiting dialogue. The dissent would have upheld the regulation as reducing the influence of certain dominant speakers in the marketplace and thereby equalizing the terms of deliberation and debate.

Consequently, the unregulated marketplace of information and data is not the appropriate standard against which to assess regulation. The dissenters in Sorrell were closer to the truth in recognizing that pharmaceutical data are a commodity that can be harvested by individuals and regulated by states for the greater good. But the dissenters too overemphasized the values of deliberation and debate. Instead, the relevant inquiry is whether the marketplace for information and data as

---

290 Sorrell v. IMS Health Inc., 131 S. Ct. 2653, 2671–72 (2011); id. at 2673–74 (Breyer, J., dissenting).
291 Id. at 2668 (majority opinion).
292 Id. at 2671.
it exists—in the case of Sorrell, the market for mining pharmaceutical data—is aligned with the normative values of transparency and commercialization. This Article has presented a portrait of these normative values, which is intended to be comprehensive, but also as a starting point for debate. Arguably, the state’s attempts to regulate the market for pharmaceutical data mining is consistent with these normative goals, as the dissent reasoned. But more to the point, the unfettered marketplace of ideas is not the appropriate ideal. Instead, the critical inquiry is whether a particular marketplace for information and data is desirable.

A piece of folk wisdom among practitioners of intellectual property and Internet law is that no judge or legislator wants to be the one responsible for “shutting down the Internet.” This folk wisdom bespeaks a precautionary principle in regulating features of the Internet and the many platforms it permits. The fear of shutting down the Internet leads to a laissez-faire view of regulation of data and information. Of course, this author does not want to be the person responsible for shutting down the Internet or any other mechanism for collecting and disseminating information. Nothing advocated in this Article will lead to that result. Instead, this Article is the starting point for a deeper understanding of the marketplace for information and data and its normative purpose. The knowledge that this understanding provides can better construct regulation of the information and data marketplace, consistent with the Constitution, general welfare, and all the marketplaces that the flow of information and data touch.

As one example of how the analysis of this Article need not threaten the Internet, the next subsection presents a proposal for data commercialization. This proposal brings together the pertinent normative concerns, the mechanisms of commercialization, and the unfortunate interpretations of the First Amendment evinced in Sorrell.

2. A Proposal for Open Source Licensing of Data

How then to respond to the issues of data commercialization raised by this Article? Shifting the First Amendment analysis, one finds a solution in Sorrell, but perhaps an unrealistic one. This Article has pointed to emerging problems posed by data commercialization—some traditional, such as privacy and security, and others novel, such as the boundaries between government management of data and private uses. The case studies above highlighted some of these issues with an argument presented for the role of competition as opposed to strong proprietary rights in promoting data commercialization. In that spirit, the Article concludes with a proposal for open source licensing of data.\footnote{293 For background on open-source licensing, see Greg R. Vetter, \textit{Commercial Free and Open Source Software: Knowledge Production, Hybrid Appropriability, and Patents}, 77 \textit{Fordham L. Rev.} 2087, 2124–26 (2009) (describing benefits of open source for proprietary software), and Greg R. Vetter, \textit{Slouching Toward Open Innovation: Free and}
largely through private initiative, would be consistent with the Sorrell decision and would permit commercialization of data within the normative boundaries set forth above.

The proposal that follows is drawn in part from the structure of Data.gov, which provides terms of use that are meant to regulate the control and dissemination of data provided by the site. Promoting certain licensing practices has been the basis for the open source movement in software and the propagation of more user-friendly copyright terms through creative commons licenses. This Article illustrates that we live in a world where data commercialization will be standard practice. This Article proposes the use of licensing terms that can effectively regulate the practices of commercializing data. This approach builds on the work of Professors Jerome Reichman and Paul Uhlir on the scientific research commons and expands them to cover data generated in the public-private model described in this Article.

The licensing terms can be implemented in several ways. The first is to have state and federal governments that make data available through websites or other fora to implement licensing terms that govern the use of the data. As the Data.gov example illustrates, such licensing already occurs, but unfortunately in incomplete ways. The ideas presented above should serve as a starting point for identifying and addressing issues raised by such licensing. Furthermore, these licensing terms should be propagated through private entities, such as social networking sites that present in their terms of use the appropriate parameters for manipulating and disseminating data generated and collected from a specific site. Competition among these private sites on these terms can serve as a basis for experimenting and determining which terms of use are desirable. Consideration of these terms of use should be part of good management practice for commercial sites.

What would constitute a model license for the regulation of data? At a minimum, the following are proposed:

- Attribution of the source of the data for any subsequent retransmission, manipulation or interpretation of the data.


Limitations on reusers to limit or deny access to the data that has been gathered or collected—similar to what currently exists on Data.gov.\textsuperscript{296}  
Warranties that data will not be revealed or used in a way that would harm the reputation or dignity interests of those who are the subject of the data gathered or collected.  
Freedom to commercialize interpretations of the data that do not violate the warranties of reputation and dignity and the access to the underlying data.  
Freedom to replicate published interpretations of the data supported through a clear statement of methodology and approaches to the transformation or manipulation of the data.  

The last model term follows practices of scientific researchers in peer-reviewed journals. Its inclusion here supports the spirit of competition and openness that ensures vitality in the usage of data. The other provisions address other issues identified in this Article, such as protection of autonomy interests and questions of fairness in terms of access to data and distributive values. While some people may be willing to sacrifice reputation and privacy to provide data that can be used for commercial gain, the model licensing terms should be designed to prevent exploitative uses of personal data.

Most importantly, these five basic model terms can support the three normative perspectives presented in this Article by allowing room for free engagement in politics and markets, while protecting values of autonomy and fairness. As the methods of commercializing data evolve, the proposed open-source-licensing model can be adapted to address the range of concerns that develop from an adaptive marketplace. The starting point is recognizing the world of data commercialization so that we will be able to more creatively and flexibly structure the regulatory response.

V. THE FUTURE OF THE MARKETPLACE OF INFORMATION AND DATA

As the Supreme Court’s analysis in its 2011 Sorrell decision indicates, the marketplace of ideas is grounded in the free flow of data and information. While the majority viewed this marketplace in laissez-faire terms, the dissent recognized that the marketplace for data is created through a network of government regulations and policies. Unfortunately, however, neither the majority nor the dissent adequately addressed either the normative framework for the marketplace of ideas or the regulatory structures that make the generation and exchange of data possible. This Article fills both of those gaps.

The normative framework for the marketplace of ideas requires a consideration of classic liberal principles with regard to freedom and rules, of autonomy with respect for individual notions of privacy, and of fairness with
respect to distribution across individuals. When applying the First Amendment as a check on regulation of data and information, courts must adequately balance these three principles, applying varying weight to each principle depending upon the context. The majority in Sorrell emphasized the concept of autonomy by deciding in favor of a laissez-faire view of the marketplace of ideas. The dissent leaned toward the classic liberal position with an emphasis on the market harms that result from a concentrated market for data arising from the market power of pharmaceutical companies in targeting doctors. Neither the majority’s nor the dissent’s analysis adequately balanced the principles of, or fully considered fairness to, patients and consumers. Such consideration may have produced a more nuanced set of decisions, as opposed to ones that turned on the dichotomy between free markets and regulation.

The majority and the dissent also failed to address the network of government regulation that makes communication of data possible. The dissent came closer to acknowledging this network, but the dissenting opinion focused largely on food and drug regulation. As set forth in this Article, however, the network of regulation is more complex, consisting of intellectual property laws as well as state and federal regulations of information provided to and used by administrative agencies. This set of regulations gives rise to a host of compelling legal problems involving the use of data—from the sale of tax assessment records to the open government initiatives under the Obama Administration. This Article describes these areas and shows that the free speech concerns raised in Sorrell may be quite extensive. Given the varied contexts for the commercialization of data, the laissez-faire approach adopted by the majority may not do justice to the range of regulatory issues and governmental interests at stake.

How government can both create and regulate the dissemination of data and information will continue to be a policy issue. Private companies see data as a resource for exploitation. At the same time, that same data is generated through the panoply of transactions and regulations that define our lives. The Sorrell majority decision leads to an unregulated market for data. Yet, the dissent’s views may become more salient once we recognize the complex web of governmental interests that inform data commercialization.