Informing and Reforming the Marketplace of Ideas; the Public-Private Model for Data Production and the First Amendment

Shubha Ghosh, University of Wisconsin - Madison
INFORMING AND REFORMING THE MARKETPLACE OF IDEAS:
The Public-Private Model for Data Production and
The First Amendment

Shubha Ghosh
Vilas Research Professor of Law
University of Wisconsin Law School

ABSTRACT

In 2011, the Supreme Court held that the First Amendment applied to the commercialization data in Sorrell v. IMS. While the case at issue dealt with state regulation of pharmacy data, the Court’s holding extends to regulation of data in many contexts from government created databases to search engines and social media sites. This Article contains a critique of the decision, emphasizing that the majority and dissent take polar opposite positions without adequately addressing the normative foundations for data regulation and the institutional arrangements within which such regulation occurs. The critique provides a normative framework for the free flow of data and information that takes into consideration classic liberal principles, autonomy principles, and fairness. This normative framework is used to analyze the regulatory structures within data commercialization occurs, including intellectual property law, state law on open records, and information access initiatives. The Article offers three examples drawn from disputes involving tax assessment data, the data transparency projects of the Obama Administration, and protection of databases in Europe, to illustrate the issues raised by the Sorrell decision. The Article ends with a set of best practices for data commercialization that takes account of the normative framework and the First Amendment as applied to data.

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, flush with data, all informative or entertaining to the new brand of digital citizens and their communities. The marketplace of ideas has become a marketplace of information and data, open for those who are able to penetrate the layers of emails, blog posts, tweets, flickrs, and 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 Supreme Court’s 2011 decision in Sorrell v IMS illustrates.

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 Broadcasting Co. v. FCC, 395 U.S. 367, 390, 89 S.Ct. 1794, 23 L.Ed.2d 371 (1969); see Abrams v. United States, 250 U.S. 616, 630, 40 S.Ct. 17, 63 L.Ed. 1173 (1919) (Holmes, J., dissenting).
This Article presents a critique of what appears to be a new approach to the regulation of commercial speech under the First Amendment. My argument identifies a misdirected normative position in how the United States Supreme Court deals with new technologies for information production and data mining. As Justice Kennedy stated in Sorrell v. IMS Health:

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 these interests, the State cannot engage in content-based discrimination to advance its own side of the debate.²

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. Specifically, under the facts of Sorrell, the Court examined the State of Vermont’s attempts to restrict the use of pharmacy data by pharmaceutical companies to direct market to prescribing physicians. While the Court is aware of the issues raised by the mining of the explosion of information, its solution is one of letting the marketplace regulate itself. Consequently, the Court struck down the regulations on data mining on First Amendment grounds. While Justices Breyer, Ginsburg, and Kagan would uphold the specific regulation on First Amendment grounds, their dissent does not address fully the normative grounds for regulating data commercialization consistent with the First Amendment. I argue that regulation of the marketplace of information is not inconsistent with freedom of speech, but 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 quoted in Sorrell “The commercial marketplace, like other spheres of our social and cultural life, provides a forum where ideas and information flourish.”³ However, this approach ignores a tension between transparency and commercialization. I identify three compelling normative positions that recognize and address this tension. I will refer to these three respectively as: (i) the classic liberal perspective (referred to in this Article as “the liberal perspective”);⁴ (ii) the autonomy perspective;⁵ and (iii) the fairness perspective.⁶

³ Id. at 2680 (citing Edenfeld v. Fane, 507 U.S. 761, 767 (1993)).
⁴ For an example of what I mean by the classic liberal perspective, see Archon Fung, Mary Graham, & David Weil, Full Disclosure: The Perils and Promise of Transparency 171-173 (2007)(describing ways in which government policy can effectively empower people as market and political actors); Alfred D. Chandler & James W. Cortada, The Information Age: Continuities and Differences, in A Nation Transformed, supra note 1 at 298-299 (describing the development of information and digital technologies as an American story, celebrating individual liberty).
⁵ For an example of what I mean by 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 face of a large bureaucracy).
⁶ For an example of what I mean by the fairness perspective, see Edith Brown Weiss, In Fairness to Future Generations258-259 (1989)(advocating intergenerational equity as a guiding principle in the management of knowledge systems, which include archiving and preservation of data about natural and cultural worlds).
Stated briefly, the liberal perspective, perhaps closest to the current Court’s jurisprudence, would conclude that there is little tension between transparency and commercialization because a fully autonomous person is one who acts in both the political and market arenas. The autonomy perspective, however, would see a potential conflict as commercialization of data leads to a confounding of personal and public spaces that can distort meanings and identity. Finally, the fairness perspective would find that 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 some light on the policy issues raised by the commercialization of data and serves in shaping how data commercialization can be organized in an effective and desirable manner.

My thesis is that contemporary debates about the commercialization of data, illustrated by the examples above, reflect an inherent tension between democratic values of transparency and accountability and market goals of wealth creation. On the one hand, democratic society demands transparency and the disinfecting influence of sunshine. This demand mandates that information and data be open and accessible. On the other hand, wealth creation through markets requires that data and information be manipulated and transformed, often in ways that distort underlying values of openness and the meanings constructed around individual quanta of data and information. This article explores this tension, demonstrating different ways in which it is identified and reconciled and concluding that some form of open source licensing, with trumps for the 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 of the First Amendment and personal privacy and dignity issues raised by the commercialization of data. Recently, Wikileaks raised issues of how transparent open government could and should be. But this controversy overshadowed more pervasive concerns about the commercialization of data. Social network sites challenged the balance between private and public information. 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 had considered that utilized information on consumption

---


8 See James Grimmelman, Saving Facebook, 94(4) Iowa L. Rev. 1137, 1149-1151 (2009)(outlining the contours of privacy protection on Facebook, and presumably other social networking sites). For a popular discussion of these issues, Jose Antonio Vargas, The Face of Facebook, The New Yorker, September 20, 2010, at 54, 63 (explaining he dynamic of information disclosure and marketing for users of Facebook).

patterns obtained from its users of email and search engine services. The company, a trend setter in the development of markets based on digital technologies and information services, expressed hesitation in compromising privacy for profit while describing innovative platforms for the exchange of data and advertising in order 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 future source of commercial value in information based economies.

The issues raised by Google’s vision of the commercialization of data goes beyond accepted Internet transactions, such as emailing, searching, and social networking. Innovations in the processing of genetic information, particularly as applied to personalized medicine and the possibility of genetic profiling, raises analogous questions of the extent to which commercialization of data is constrained by social concerns over privacy and individual autonomy and legal concerns over the scope of property rights. The broad question raised by both Internet transactions and genetic information is to what extent can information generated by individual persons be the basis for business models for the collection, aggregation, and processing of information.

This question, however, extends beyond the realm of new information technologies. For example, an issue in the 2010 Census was the recording of same sex couples as married 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. The Google memo, personalized medicine, and the Census example, arising from very different technological, legal, and social environments, raise a common question: how can data and information be used in a liberal, democratic society based on a market economy? Is it possible to reconcile values of transparency and openness with those of commercial gain and ownership? This paper addresses these questions as matters of legal entitlements and institutions and the construction of social meaning.

Government policy attempts to navigate the tension between transparency and commercialization through its dissemination of data, but often in misdirected ways. For example, data.gov is a web site created by the Obama Administration which makes accessible a range of government databases, covering such topics as atmospheric data, employment statistics, and

---

10 Id.
11 Id.
13 See “A Sea of Sensors,” The Economist, November 6, 2010, at S6-S7 (describing information gathering sensors and their relevance for commercialization, pricing, and marketing); “London: Turning Access Into Apps,” Time, January 17, 2011, at 6 (how state collected information serves as the basis for iPhone applications).
government expenditures.\textsuperscript{16} The terms of use for the site do not prohibit the distribution of data or the commercialization of the data, as long as the data is distributed without restrictions on reuse of data. Furthermore, the terms make a distinction between data and information. Data is defined as “values or sets of values representing a specific concept or concepts.”\textsuperscript{17} According to the site, data becomes information when combined to extract meaning and put data in context. Put another way, data is raw or uncooked information while 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 would not prohibit 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 proprietary domain where market transactions can be purely private and value can be appropriated (information).

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 an elusive, and perhaps non-existent one. 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. The Supreme Court in Sorrell leaves the matter to the marketplace by limiting the government’s ability to regulate the commercialization of information and data. I show 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, described briefly above, that informs both the First Amendment and the purposes of a marketplace for information and data.

The structure of this Article is as follows. Section Two presents an analysis of the Sorrell decision, placing it in the context of the First Amendment jurisprudence and a brief history of information and data commercialization. Section Three builds on the economic realities by turning to the legal structure of data markets, rooted in intellectual property and other information-regulating laws, such as privacy and security. Section Four presents the critique of what I describe as the emerging commercial speech paradigm. Section Five summarizes and concludes.

II. Sorrell and the Marketplace of Ideas, Information, and Data

The Supreme Court’s 2011 decision in \textit{Sorrell v. IMS} is a touchstone for understanding the normative issues raised by the commercialization of data. Unfortunately both the majority and dissenting decisions did a poor job of addressing the broader normative and policy concerns. 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 with the remaining sections building on my analysis of the opinion to construct a deeper jurisprudence of data.

\textsuperscript{17} See FAQ, available at http://www.data.gov/faq (accessed on February 9, 2011).
At issue in Sorrell were state regulations restricting the use of pharmaceutical data for the purposes of targeting advertising and promotions at prescribing doctors. This practice called “detailing” is used by pharmaceutical companies with on-patent pharmaceuticals who have their sales representatives contact physicians directly in order to promote their products. About half a dozen states enacted regulations that prohibited this practice by preventing the sale of pharmaceutical data by pharmacies and insurers and the use of such data for marketing purposes by pharmacies, insurers, and pharmaceutical companies and marketers. Typically the data were collected by private companies, referred to as data miners, who aggregate and package the date for purchase by the pharmaceutical companies.

The states enacted these restrictions on the practice of detailing in order to contain the costs of health care and prescription drugs as detailing was a practice of the large pharmaceutical company rather than the generic drug companies. Furthermore, the states sought to protect the privacy of physicians whose information were contained in the pharmaceutical data. Relevant privacy statutes, such as HIPPAA, protected the patient’s privacy, but not the doctor’s. The prohibitions on detailing were meant to protect doctors from intrusive advertising and marketers. Claiming violations of their First Amendment rights, a group of data mining and pharmaceutical companies challenged the statutes. The Maine and New Hampshire statutes were held constitutional by the First Circuit in IMS v. Mills and IMS v Ayotte, but the Second Circuit struck down Vermont’s statutes in IMS v. Sorrell. This circuit split lead to a grant of the certiori petition by the US Supreme Court in Sorrell.

In a six to three decision, the US Supreme Court affirmed the Second Circuit, finding that the Vermont prohibitions on sale and use of data were a content and speaker based restriction on speech. The dissent would uphold the regulations as “a lawful governmental effort to regulate a commercial enterprise” and 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 regulation, and (c) the effect of the regulation on First Amendment values. As I argue later in this Article, neither address the core economic, social, and political values raised by data commercialization beyond the narrow confines of freedom of speech. But before I address this broader claim, let me 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 such core categories as political speech. A government restriction on commercial speech is reviewed by a court using an intermediate standard of review, requiring 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 by attempting to balance the communicative channels to prescribing doctors that the state had deemed to be biased towards large pharmaceutical companies.18 The governmental goals of reducing health care costs and protecting citizens, specifically doctors, from intrusive advertising, satisfied the requirements of intermediate scrutiny.

18 See supra note 2 at 1594.
The majority construed that speech at issue differently.\textsuperscript{19} Although the opinion does recognize that commercial speech is at issue, the majority found the government failed to meet its burden under the intermediate standard of review. Justice Kennedy, writing for six justices, however, goes further to reason that strict scrutiny would apply to the regulation. Initially describing the speech at issue as “speech in aid of pharmaceutical marketing,”\textsuperscript{20} Justice Kennedy concluded that the state regulation was based on the content of speech and the identity of the speaker.

First of all, the speech was deemed to be content based because it targeted speech that promoted the marketing of pharmaceuticals ostensibly to promote or subsidize speech to promote the marketing of generics and low cost substitutes. In the advertising marketplace, the state had chosen sides, according to the majority opinion, attempting to shut down one channel of communication in order to remove the advantage that large pharmaceutical companies had in acquiring and analyzing marketing data. The regulation was not the broad based commercial regulation that had some incidental effect on speech. Nor was it targeted at specific types of commercial speech, such as attorney advertising, across the board. Instead, the restriction on detailed favored one advertiser over another by prohibiting a specific marketing practice. This regulatory strategy made the statute a content based restriction on speech, subject to strict scrutiny. The state has 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 speech was deemed to be based on the identity of the speaker. The target of the restriction was pharmaceutical companies that used the data purchased from detailers to communicate with doctors to affect their prescription practices. Although the State of Vermont, in a seemingly last ditch effort, contended that the restriction was aimed at sale and use by any entity that obtained the data from the pharmacies, the majority was not convinced. The opinion commented on this last ditch effort as a switch made during oral arguments from an earlier contention that the restriction was aimed only at pharmaceutical companies. But even if the Court accepted the State’s later contention, the majority opinion emphasizes exceptions in the Vermont statute allowing pharmacies to sell the data to universities or other entities for research purposes. The statute, the majority concluded, was aimed not only at specific content, but at specific speakers. Strict scrutiny would apply on either ground.

Justice Kennedy’s opinion from its start construes the speech narrowly as “speech in aid of pharmaceutical marketing.” The majority rejected the argument persuasive to the First Circuit (which upheld similar statutes against First Amendment challenges) that the speech at issue as a mere commodity that was bought and sold rather than communicative expression. The Court rejected this characterization as follows:

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

\textsuperscript{19} See supra note 2 at 1561.  
\textsuperscript{20} Id.
First Amendment purposes.\textsuperscript{21}

The data at issue were not mere commodities that could be regulated by the state but valuable inputs to decision making and the formulation of knowledge and opinions. The Court is saying that in a marketplace for ideas, consumers need access to such data. While the state would have an interest in regulating such data as speech if false, the majority found the speech to be useful in physician decision making. Therefore, the state could not simply restrict the speech just 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 pure commercial speech, the communications at issue under the regulation, the transfer of data, its processing, 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 differ on how that marketplace operates when dissemination of information is its object.

B. The Nature of the Regulation

The majority and dissent also diverge in their characterizations of the regulation. As discussed above, the majority views the regulation as based on the content of the speech and the identity of the speaker. The dissent by contrast views the regulation as part of a comprehensive scheme to regulate pharmaceuticals and protect the interests of patients.\textsuperscript{22} Not surprisingly, the two contrasting depictions lead to opposing assessments of the First Amendment claim.

As the majority admits, all government regulation has some effect on speech.\textsuperscript{23} To take the majority’s own example, laws against employment discrimination or sexual harassment affect what is said and how in advertisements for jobs and in everyday office conversation. But such effects do not transgress the First Amendment because they are deemed incidental. According to the majority, the Vermont statute directly regulated speech by restricting use of certain data that made possible the communications between marketers and physicians. Regulation of speech was core to the Vermont statutory scheme and the goals of protecting privacy and controlling health care costs were deemed incidental. The dissent on the other hand sees the effects on speech as incidental to the legitimate goals of the state in regulating health care and the pharmaceutical industry.

One argument critical to the dissent’s position is that the prescription data at issue exists because of government regulation.\textsuperscript{24} Federal food and drug regulation requires prescription for certain medications. These regulations require the prescriptions to be of a particular form and to be processed in a particular way. Absent this requirement, the physician-patient relationship

\textsuperscript{21} Id. at 1571.  
\textsuperscript{22} Id. at 1595.  
\textsuperscript{23} Id. at 1560.  
\textsuperscript{24} Id. at 1596.
would largely be a matter of pure contract. Physicians might recommend certain drugs as treatment and may even dispense them directly. The structure of pharmaceutical distribution would be very different absent government regulation. Since the pharmaceutical data itself is a creature of regulation, the dissent concluded that the government has wide latitude in regulating such data. This characterization of the data at issue contrasts with the majority’s view that pharmaceutical data is private data because it is in the hand of a private entity, the pharmacies. By emphasizing the private possession of the data, the majority distinguishes 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.

The majority’s view of the data emphasizes the possessory interest and ignores the governmental regulations that effectively created the data. In short, the majority’s analysis pays short shrift to the regulatory environment surrounding the pharmaceutical data and reduces its ownership to private hands, free from regulation, 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 assumes that since government regulation created the pharmaceutical data, the government owns the data and can do what it wants with it. That 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.

Both the majority and the dissent characterize the data too simplistically, one overemphasizing its private nature and the other its public roots, in reaching its conclusions about the 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 Three of this Article provides a corrective, but the majority and dissent in Sorrell each illustrate the problems that arise from not fully understanding the marketplace for data and information.

C. The Marketplace of Ideas

The last section ended by pointing to some of the normative deficiencies in the majority and dissenting opinions. Nowhere is this more obvious than in the contrasting view of the marketplace of ideas that informs the respective understanding of the First Amendment by the majority and dissent. Critics of First Amendment jurisprudence have pointed to the deficiencies of the marketplace of ideas as a paradigm for adjudicated 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, what could be captured by the phrase “free play of ideas,” de-emphasizing the potentially misleading use of the word trade. The majority and dissent both rely on the marketplace of ideas, but with different implications. Consequently, I will engage with the concept of marketplace of ideas here 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 the participants in the marketplace. Under this view, any attempt by the government to favor one side or to tilt the marketplace in one direction would be inconsistent with the First Amendment. The suspicion towards content based regulation as well as speech regulation based on the identity of the speaker stems from the government attempting to intervene into the free trade in ideas made possible by the marketplace.

In the case of Sorrell, the majority strikes down the Vermont statute precisely because it chose sides in the marketplace and attempted to tilt it in favor of some speakers at the expense of others. Although the state expressed concern that large pharmaceutical companies were able to gain access to physicians while smaller generic companies could not, the majority attributed this difference in access to the persuasiveness of the pharmaceutical companies. The First Amendment prohibits the government from punishing or prohibiting this persuasive success simply because it does not like the result unless there is a strong justification. Neither protection of physician privacy nor health care cost containment provided this justification. Doctors, as the majority points out, could simply ignore the marketers. And health care costs could be contained through means that do not interfere with speech. In fact, communication about products is a main way in which efficient and accurate pricing can occur. The state cannot regulate price through restricting advertising.

In all respects, the majority adopts a laissez-faire view of the marketplace of ideas. A contrasting view is the foundation for the dissent’s First Amendment analysis. For the dissent, the marketplace of ideas is not a perfectly functioning market. The government needs to intervene to protect against typical market deficiencies like concentration of resources and unequal access. The dissent’s view is the marketplace of ideas requires a level playing field and government intervention can help to secure such levelness. 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 resources to extract this resource and use it in communicating with physicians to better market their products. Statutes, such as the one enacted by Vermont, serve to limit the advantage gained by pharmaceutical companies and permit generics to compete more effectively. Instead of undercutting the marketplace of ideas, the Vermont statute reinforces it. Therefore, the dissent concludes, the statute is consistent with the First Amendment.

The Sorrell case is an important case for the constitutional standard by which to gauge regulation of the market for information and data. Both the majority and dissent construe this marketplace in terms of the marketplace of ideas, the workhorse for First Amendment analysis. But as I have pointed out, both majority and dissent are deficient in their normative assessment of the regulation and the free speech values at issue. The two sides are divided in their assessments of the marketplace of ideas, but neither sees beyond this marketplace as the gauge for understanding how information and data are generated and disseminated. There are values beyond the marketplace of ideas. In the next section, I flesh out the richer legal and normative

25 Id. at 1583.
26 Id. at 1598.
structure of the marketplace for information and data. In Section IV, I examine other examples of such marketplaces and return to the case of Sorrell in light of the this richer legal and normative structure.

III. The Legal and Normative Structure of Information Markets

This section elaborates upon my point that the Court failed 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 a failure of litigants and commentators to fully describe the legal structure of information markets and its 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 can be created, distributed, and consumed. An account of the legal structure of 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 those 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 structure with data.gov being just one example of the role of the government in creating, disseminating, and consuming data. 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 I will show in this section and in the case examples of Section Four, the government is a key player in such situations as the commercialization of real estate data and census data.

Understanding the legal structure is a step towards assessing data commercialization against a normative framework. As explained in Section One of this Article, there are three normative perspectives one can take towards the practices of data commercialization:

1. the liberal perspective;
2. the autonomy perspective;
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 data on same sex couples from the Census for advertising purposes is consistent with the liberal perspective because the use of such data recognizes the
economic value and autonomy of the 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 a different 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, allowing market recognition of same sex couples while denying political or legal recognition. Legal institutions should respond to this discrepancy accordingly by giving full weight to the autonomy of individuals.

Finally, the fairness perspective would consider the differing bargaining powers of actors in the political and market realm and scrutinize the commercialization of data from a distributive justice perspective. With respect to the Census example, the fairness perspective would treat the collection and use of the data as a potential exploitation of same sex couples and would support a ban on the collection of the data. This conclusion 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 and perhaps even a step towards 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. As I show, the actual legal structure governing data commercialization represents a compromise among these three normative perspectives. In this way, data commercialization is made possible while acknowledging competing normative perspectives. The analysis presented in Section Three will provide the foundation for the licensing model for data presented in Section Four, 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 their restrictions on how each can be sold or shared? These are the questions addressed in the rest of this subsection.

1. Intellectual Property Regimes

Four bodies of intellectual property law govern the ownership of data. I discuss them in order of doctrinal and normative complexity; copyright, first, then patent, trade secret, and common law misappropriation in that order. I end this discussion with a short analysis of overarching First Amendment principles.
Copyright. The principal United States case directly addressing the question of ownership of data is the 1991 Supreme Court decision in *Feist v. Rural Telephone Company*. At issue in the Feist case was the copying of entries in a telephone directory compiled and issued by the Rural Telephone Company in Kansas. 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 (in 1991 that meant manual and visual searches rather than digital ones) the entire state. Ruling on the claim that Feist had infringed the Company’s copyrights in the directories, the Supreme Court emphasized two legal principles. First, one requirement for copyright protection is originality, which means that the purported author of the work was the original source for the protected material and that the author imbued the material with his or her own creativity. 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. Instead, copyright protection extends to the creative selection, arrangement, and coordination of the facts and data by the author making the compilation. The Company was deemed not to have copyright in the alphabetical arrangement of names and other data in the phone directory, and Feist was free to copy that list of names and other data. The Court emphasized in its decision that copyright protection for database was thin, extending only to the particular selection, arrangement, and coordination of the facts and data, creatively chosen by the author.

The synopsis of the *Feist* decision conflates the terms facts and data. The Court in its opinion refers to facts and makes references on occasion to data. The Court’s rationale, however, applies 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 database and independent of the compiler. Therefore, data can 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. The data.gov website 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, but 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 Days” contains both qualitative and quantitative facts. The quantitative fact would be data according to the working definition of data. This distinction may not be all the satisfactory because “values” can be qualitative as well. For this reason, I will leave others to make meaningful distinctions between facts and data. In this Article, I will use the term data synonymous with facts, and particularly in discussing copyright law.

The example of Phileas Fogg illustrates a more important 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

28 See supra note 27 at 348.
29 See supra note 27 at 349.
30 See supra note 27 at 346 (referring to raw data as “wholly factual information not accompanied by any original written expression”).
First, this datum is expression and as expression it 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. An additional requirement for copyright protection is that the expression be a recognized work of authorship, such as the compilation at issue in *Feist.* Other recognized works of authorship include literary works, musical works, and audiovisual works. A single sentence can constitute a literary work although the Copyright Office through its regulations has excluded short phrases, such as advertising jingles, from copyright protection in order to draw a distinction between copyright and trademark laws. A longer work, like this paragraph or a novel, containing the sentence would be a literary work, and the copying of the sentence would provide evidence of substantial similarity, the legal test for copyright infringement. In this way, copyright protection extends to the way in which data is expressed.

Second, copyright law extends only to the expression, and not to the idea being expressed. The data contained in a work of authorship would not be subject to copyright protection, as the Supreme Court made clear in *Feist.* Therefore, anyone can report that Phileas Fogg went around the world in seventy-nine days. Copyright law intrudes when the specific expression of that datum is copied. The ruling in *Feist* is just a specific example of this broader notion in copyright law. 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. 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 extend to building block components of the work, such as a letter, a note, or data.

Third, the challenging question is the treatment of created or constructed data. In the case of the Phileas Fogg example, the datum was initially created by Jules Verne 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. This principle pushes in favor of protection for created data. On the other hand, data is meant to be openly 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 non-fictional 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

---

33 See 17 USC 102(b).
34 See supra note 27 at 347.
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, particular with respect to the creation of items like stock indices or sports statistic. 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 in order to protect accessibility, and the latter can in order 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 creator of the index or the sports statistic to 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, like a database, a way of processing the underlying data.

The approach under Feist contrasts with that adopted by the European Database Directive in 1996. The Directive was enacted in response to concerns by members of the database industry that the thin copyright protection under Feist was not adequate to protect the economic interests of legal and financial database owners. The Directive extends protection for fifteen years to creators of databases which are the product of “substantial investment in obtaining, verifying, or presenting the contents of the database.” Database owners have the right to prevent imitation or alteration of the database and extraction of the data from the database. I discuss the Database Directive in more detail in Section Four of this Article. What is important at this stage is to identify the differences between the Directive and the approach pursuant to Feist. First, the Database Directive extends protection to the underlying data itself, to the extent the data is extracted from the database. Second, the Directive bases protection of databases on

36 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.”)
38 Id.
39 See discussion, supra at text accompanying notes 12-13.
41 See Kregos, supra note 37 at 705; Feist, supra note 31 at 348-349.
43 Id. at art. 55.
44 Id.
a “substantial investment” by the creator rather than any standard of creativity.45 Arguably, an alphabetic listing of names as in Feist would be protected if the compilation of such a list required substantial investment of time and money. Under Feist, a mere alphabetic compilation would never be protected by copyright law. The Directive therefore expands the scope of database protection and rights of the database creator. The United States had considered similar legislation in the 1990’s, but opposition from those seeking to protect access to data blocked the passage of analogous laws.46

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 web sites 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 expression without also taking the data. For example, if a post says “I partied all night yesterday,” the data contained in the sentence cannot be separated from the sentence itself.47 Therefore, under the merger doctrine, the sentence itself loses copyright protection. For all 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 data. 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 synthesis 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 the 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 extraction of data.48 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

45 Id.
47 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).
48 See Reichman & Samuelson, supra note 46 at 66-67.
can constitute a database. But what about the web site itself? To what extent is all of Facebook a database? These questions have not been directly addressed, but they raise the possibility of a meta-database or a database that consists of other databases.\textsuperscript{49} The difficult possibility is that 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{50} In this way, the owner of the meta-database may end up effectively obtaining ownership of the data and limiting access. 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 web site itself will stake a claim in the aggregate of the databases. Conflicting and competing ownership interests may mitigate the ability of the meta-database owner to restrict access and thereby continue to keep data free from copyright protection.

**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 database uncovers several hundred patents pertaining to the processing of data and the manipulation and use of databases.\textsuperscript{51} Patents, owned by companies like Google, Microsoft, Yahoo, and several smaller Internet based and computer outfits, cover search of databases, extraction of data, coordination and linking of databases, and other ways in which data can be arranged, coordinated, selected, and transformed.\textsuperscript{52} Professor Pamela Samuelson has argued that patents on data interfaces impede the interoperability of data and information and transfer mechanisms across systems, applications, and platforms.\textsuperscript{53} 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 composition of matter.\textsuperscript{54} Case law establishes three exclusions from the definition of invention: abstract ideas, laws of nature, and natural phenomena.\textsuperscript{55} Data itself would not be patentable since it does not fall into any of these four categories of invention. Processed data in the form of information or a

\textsuperscript{49} See id. at 114-115.
\textsuperscript{50} See id. at 103-104.
\textsuperscript{51} A search of patents in the USPTO online database uncovered 4343 patents which had the word “data” or “database” in the title and in the abstract. The search was conducted on February 10, 2011.
\textsuperscript{52} See, e.g., “Method and Apparatus for Exchanging Data with a Database,” Patent # 7877417 (January 25, 2011)(patent assigned to Microsoft covering invention that allows for extraction of data from database of any structure); “Searching Structured Geographic Data,” Patent # 7836085 (November 16, 2010)(patent assigned to Google for searching and extracting data from multiple datasets); “Expanding a query to include terms associated through visual content,” Patent # 7882124 (February 1, 2011)(patent assigned to Yahoo! for searching databases containing visual content).
\textsuperscript{54} 35 USC 101.
database would also not fall under patentable subject matter. However, tools to extract, aggregate, manipulate, or process data would be considered as a patentable invention. A general purpose machine, such as a computer, can be patented. Similarly, a special purpose machine, such as a heart monitor, that collected, extracted or analyzed special types of data, such as health indicators, can be patented. Furthermore, processes for manipulating or transforming data that are not necessarily tied to or operationalized through a machine can be patented. So methods for search or for data extrapolation can be patented. A patent is obtained after an application and review process with the United States Patent and Trademark Office. The administrative process checks to see if the invention is novel, nonobvious, useful, and a patentable invention. Once granted, the patent allows the owner to exclude others from using, making, selling, or importing the patented invention for a period of twenty years from the date of application.

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. A challenger to a patent would have to 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. Novelty and nonobviousness entails an analysis of the prior art leading up to the invention while usefulness and disclosure requires a consideration of the patent application and the nature of the invention. As 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.

As for novelty or nonobviousness, the validity question would rest on the prior art in the relevant field within which the data related invention falls. The novelty inquiry rests on a fairly broad understanding of the prior art; it can be from any field. The nonobviousness inquiry rests on a narrower scope of the prior art; it can come from an analogous field to the invention at issue. For the most part, the prior art for data related inventions will be from the fields of computer science or electrical engineering. Each of these fields will present the state of knowledge on electrical processes, computer flow charting, and data processing that will aid in assessing the invention. The novelty analysis will focus on whether an exactly identical invention already existed in the prior art. The nonobviousness analysis will focus on whether the differences between the patented invention and the what was known in the prior art would be

---

57 See In re Alappat, 33 F.3d 1526, 1545 (Fed. Cir. 2003)
58 Id.
59 See In re Alappat, 33 F.3d 1526 (Fed. Cir. 1994); State Street Bank & Trust Co. v. Signature Financial Group, Inc., 149 F.3d 1368 (Fed. Cir. 1998).
60 35 USC 131-13.
61 35 USC 101-103.
62 35 USC 271.
63 35 USC 282.
64 See e.g., In re Bilski, 545 F.3d 943, 976-977 (2008)(Judge Newman’s dissent identifying issues of novelty and nonobviousness in addition to patentable subject matter in addressing patenting of processes).
65 35 USC 102-103.
66 See, e.g., In re Lowry, 32 F.3d 1579 (Fed. Cir. 1994)(patent for invention involving data processing system).
obvious to a person having ordinary skill in the art.\textsuperscript{66} In other words, do the differences constitute trivial variations or substantive progress in the field? On this last inquiry, data related patents may be subject to vulnerability as there may be very few new methods of processing data that are truly innovative. To emphasize once again, 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.\textsuperscript{69} At issue in *Bilski* was what types of processes would constitute a patentable invention.\textsuperscript{70} 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, meaning that the disclosure merely described a series of steps, but did not disclose any operational or engineering steps.\textsuperscript{71} The Board of Patent Appeals and Interferences, the first level of review within the USPTO, affirmed the rejection but based the rejection on the grounds that the invention was an abstract idea, as opposed to a concrete application.\textsuperscript{72}

The United States Court of Appeals for the Federal Circuit reviewed the rejection and upheld it but on substantially different grounds. Reviewing the Supreme Court precedent on the patentability of processes, particular software related processes, the Federal Circuit held that not all processes are patentable.\textsuperscript{73} The court held that in order for a process to be patentable it either had to be machine based or constitute a transformation.\textsuperscript{74} The difficult part of this “machine or transformation” test is discerning when a process constitutes a transformation.\textsuperscript{75} Partly, a transformation involves something physical, such as a change from one physical state to another as in a chemical or biological process.\textsuperscript{76} 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. In case of the hedging 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.\textsuperscript{77} The process disclosed was purely mental or cognitive.\textsuperscript{78}

The Supreme Court granted certiori on Bilski’s appeal from the Federal Circuit and affirmed the rejection, but not the court’s reasoning. The Supreme Court held that processes in general were patentable unless they constituted an abstract idea, a law of nature, or natural

\textsuperscript{69} Bilski v. Kappos, 130 S.Ct. 3218 (2010).
\textsuperscript{70} Id. at 3221.
\textsuperscript{71} Id.
\textsuperscript{72} Id.
\textsuperscript{73} See supra note 64 at 952-953.
\textsuperscript{74} Id. at 956.
\textsuperscript{75} Id. at 957.
\textsuperscript{76} Id.
\textsuperscript{77} Id. at 958.
\textsuperscript{78} Id. See, also, In re Comiskey, 554 F.3d 967 (Fed. Cir. 2009)(rejection of arbitration method patent as invalid subject matter because the method was an abstract idea and mental step).
phenomena. The machine or transformation test was not the sole test to determine a process would qualify as an invention, but was one of many possible tests. Finally, the Supreme Court affirmed the rejection on the ground that the hedging method was an abstract idea and therefore excluded from the meaning of an invention.

In light of Bilski, data related patents may face some vulnerabilities. The Bilski opinion is an opaque one, revealing certain fault lines in the Roberts court. The unanimous decision speaks towards 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 found an exception for 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. It is unlikely, however, that the Supreme Court would recognize such a broad exclusion, especially now that Justice Stevens has retired. A more realistic threat to data related patents would be the interpretation of Bilski by the USPTO and the Federal Circuit. The Supreme Court recognizes limitations on the meaning of invention and did endorse the machine or transformation test as one test for patent validity. Subsequent to the Bilski decision, the USPTO issued proposed rules for analyzing patentable subject matter. Future developments will show how far these limitations will be taken. What is the case, however, is that patents serve as an important type of intellectual property for the protection of methods of data processing which complement the protection to information and databases protected by copyright.

Trade Secret. Inventions related to data and databases, information, and databases, and even data itself can be protected as a trade secret. Complementing both patent and copyright, trade secret law provides a comprehensive form of intellectual property, grounded mostly at the state level and through criminal enforcement by the Department of Justice at the federal level. 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 protection for many reasons including access to federal courts and the absence of strong defenses to infringement like reverse engineering, which is a critical feature of trade secret law. Stated succinctly, trade secret offers broader coverage of potentially perpetual duration but subject to stronger protections for alleged infringers, as compared to patent and copyright.

79 Bilski, supra note 69 at 3222.
80 Id. at 3228.
81 Id. at 3226.
82 Id.
83 Id. at 3229.
84 Id.
85 Id. at 3227. (Court hesitant to adopt categorical rules).
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. 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 is broad-ranging and includes processes, formulas, and data. For example, customer lists are protected as a trade secret in all jurisdictions and 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 wide subject matter, including data, there are several limitations on the role of trade secret in commercializing data. First of all, what counts as a trade secret depends on 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 does not protect privacy interests in data but only commercial 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. Owners and operators of social networking sites would not be able to protect such readily ascertainable information as a trade secret. The exclusion would extend to data contained in items posted by users. If such posts are public and readily searchable, then their data content would be excluded from trade secret protection.

Third, trade secret protection requires affirmative steps by the purported trade secret owner who must take reasonable steps to maintain the secrecy of the protected information. Such reasonable steps could include technological measures (for example, encryption or password protection) or contractual measures such as nondisclosure agreements or covenant not to compete. Reasonable steps could also include the creation of special relationships which would impose a duty of confidentiality on the parties. This requirement makes the protection of trade secret particularly difficult for the use of trade secret law in the commercialization of

---

89 The Uniform Trade Secret Act (UTSA) defines a trade secret as:
[I]nformation, 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. UTSA § 1(4) (1985).
91 See, e.g., Dow Chemical Co. v. United States, 476 US 227, 239 (1985)(trade secret not relevant to reasonable expectation of privacy analysis);
92 See UTSA, supra note 89.
93 See UTSA, supra note 89.
data. Often data will be created by many users independent of any relationship with the trade secret owner. Therefore, it will 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 imposed confidentiality requirements on the data and information generated on the site may be difficult to enforce (and would face the same technical and social challenges as protection for personal privacy). While reasonable steps are a loose requirement and may in some instances be a formal one, the vulnerability of data generated by multiple users would be grounds for questioning whether this requirement for trade secret protection has been met.

In theory, trade secret law can protect data. In practice, however, such protection can be difficult to obtain and sustain. As a consequence, trade secret protection may be limited to protection for information and databases as well as for methods of processing and manipulating data, the same subject matter as copyright and patent. In this way, trade secret law is an important complement to federal intellectual property protection and may aid in data commercialization through longer duration of protection.

**Misappropriation.** Despite its uncertain roots, the common law tort of misappropriation provides a quasi-property right for certain types of data and information. Traced to the Supreme Court’s decision in *INS v. Associated Press*, the tort of misappropriation allows the aggregator, collector, or disseminator of certain types of data or information to prevent a competitor from improperly accessing and using the data or information for a limited period of time. 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 claimant. Courts typically focus on whether a party has simply copied someone else’s data or information or has added value in how the data or information is generated or used. In this way, the common law tort of misappropriation extends trade secret law to protect items that would not qualify as a trade secret.

Timing is key to the tort of misappropriation. The data or information protection must fall into the category of “hot news,” in other words data or information 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 itself that was gathered from the battlefields of Europe. In other cases, hot news has included sports scores and market prices. Unlike trade secret law, the tort of misappropriation is designed to protect the labor of the creator in collecting, aggregating, and manipulating data.

---

95 248 U.S. 214 (1918).
96 See National Basketball Association v. Motorola, 105 F.3d 841 (2nd Cir. 1997)(presented a multifactor test as basis for valid misappropriation claim, where factors included timeliness of information, effort expended to obtaining information, free riding in copying information, and competition between plaintiff and defendant).
97 Id. at 845 (discussion requirement of free riding).
99 See supra note 95 at 231.
100 See Motorola, supra note 125 at 854. For a successful misappropriation claim involving stock indices, see Board of Trade v. Dow Jones & Co., 456 N.E.2d 84 (1983)
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. 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 niche left open by copyright, patent, and trade secret.

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 is consistent with the liberal perspective. Allowing some property protection for information, processes, and secrets promotes participation in both the market and political spheres. Furthermore, the construction of data and information as property makes no distinction between the political and economic uses of information. In addition, 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 would interpret 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, which permits 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. This middle ground will be explored more in the section below on sharing and selling data. One potential conflict between the liberal and autonomy perspectives is that this middle ground simply may not exist for the liberal perspective as all data is divided in that which is raw and open and that which is refined and proprietary.

Finally, the fairness perspective would cast this ownership structure in terms of competing sources of power, both political and economic. Allowing data to be openly accessible creates a zone free of the exercise of commercial and political power. Data transparency serves to ensure equal access and a commons free of 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

---

101 See supra note 71 and discussion infra at Section IV.A.3.
and political agents. On this point, the fairness perspective would ask more of the ownership structure than I have described at this point. It would call for more internal limiting doctrines (such as fair use and the first sale doctrine)\(^{102}\) as well as external limiting doctrines (such as laws protecting privacy, unfair competition, and misuse of personal information).\(^{103}\) 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 US Constitution and comparable protections in other jurisdictions.\(^{104}\) 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.\(^{105}\) The autonomy perspective would see the First Amendment limited by concerns for privacy and security in the middle ground that arises between raw data and highly refined data.\(^{106}\) In the two tails, however, the First Amendment would serve as a complement to the ownership structure, much as with the liberal perspective. Finally, the fairness perspective would use the First Amendment as a limiting principle to protect against power imbalances in the various spheres in which data exists.\(^{107}\) 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.\(^{108}\)

In short, the ownership structure over data is both descriptively and normatively complex. Each of the normative perspectives plays a role in 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

\(^{102}\) 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 serving to protect access).


\(^{104}\) See, e.g., IMS Health Inc. v. Sorrell, 630 F.3d 263 (2nd Cir. 2010), cert granted by Sorrel v. IMS Health Inc., 131 S. Ct. 857 (2011)(First Amendment challenge to state statute prohibiting the sale, license or exchange of pharmaceutical data).

\(^{105}\) As examples of this view of the First Amendment, see Eldred v. Ashcroft, 537 U.S. 183 (2003)(Court finding narrow First Amendment limits on copyright law); see Sorrel, supra note 133 at 274 (affirming First Amendment claim to allow for commercialization of data).

\(^{106}\) 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).


\(^{108}\) See, e.g, Houchins, supra note 107 at 12 (discussing interests in protecting inmate population from news reporting and gathering).
particular context of government shapes the particular application of the intellectual property rules. For example, federal government bodies cannot own copyright in legal materials under the United States Copyright Act unless the materials are created by an independent contractor who transfers copyright to the government. However, the treatment of databases created by the federal government would be different from legal materials, such as statutes, regulations, and case opinions. Such databases would be subject to copyright protection. There are no statutory restrictions on ownership of copyright by state governments although some jurisdictions have created restrictions for state legislative and regulatory materials. As far as patent law, governments have not obtained patents in data or information related inventions although they could. However, governments do own trade secrets in data and information, which has been the source of controversy for proponents of open government. Finally, misappropriation would exist as a cause of action for governmental entities although no reported cases illustrate such claims.

Government owned data and databases pose challenges for public accountability and access. All fifty states have enacted an open records act, 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 stated 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 in 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. The current version of the Open Records Act has been the subject of controversy and will serve as a case study in Section Four.

---

109 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, 663-668 (2004)(analyzing importance of excluding statutes, regulations, and case law from copyright protection from a democratic governance perspective).
113 Kidwell, supra note 112 at 1028.
114 Kidwell, supra note 112 at 1030.
115 Kidwell, supra note 112 at 1028.
An analysis of the current version of the Wisconsin 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:

it 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….117

The provision concludes by creating a “presumption of complete public access” with access being denied “only in an exceptional case.”118

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.119 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 as well as they arise in governmental proceedings. The key concept, however, is neither information nor data, but a record.120 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.121

The authority referred to in the definition is a governmental authority. As we will see in the case study in Section Four, this authority can include a non-governmental authority to whom power has been delegated by a governmental authority. The definition speaks broadly in terms of the medium that would constitute a record and the potential types of information that would constitute a record. However, the definition excludes copyrighted and patented materials and published materials that are on sale or available in a public library from the definition of a record.122 Also excluded are drafts, notes, and other preliminary materials used to create a record.123 These exclusions protect certain property and economic interests as well as the ability

118 Id.
119 See definition of record, infra.
120 See Kidwell, supra note 112 at 1026-1027 (drawing distinction between record and copyrighted material).
122 Id.
123 Id.
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 trade secrets, computer data, and proprietary data from a record request. 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 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 as a request. There is, however, authority to exclude these items from a record request.

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 create an official recording of governmental acts that are subject to the governmental stewardship and to citizen requests for access. 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 or 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 in order to allow access of sorts to the records and data created. The open records act 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 Four will explore these differences in greater detail in the context of case studies.

Government ownership of data has three different dimensions, each reflecting the normative complexity of the legal structure of data ownership: (1) the role of government as 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 would consider strong overlap between the market and political realms. The autonomy perspective would see conflict, however, especially if the government’s role as regulator feeds into its role as market participant, perhaps through the use of data obtained through the regulatory process by state owned enterprises. Finally, the fairness perspective would be concerned with the undue influence of the state in the private realm through its coercive and potentially monopoly power. As the case studies in Section Four 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 would consider the archiving function 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, and this is a point that will be examined more in Section Four in a case study, the web site data.gov exemplifies the liberal perspective as data accessibility provides a way to both gain knowledge useful in the political realm with minimal limits on commercialization. But both the autonomy perspective and fairness perspective would support a cautionary take 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 would draw our attention to power imbalances in how data is recorded and stored. The rich can have records expunged or controlled while the poor may not have this capability.

Finally, the normative tensions are apparent in the transparency and openness provided by both open records acts and web sites like data.gov. The liberal perspective would herald transparency, echoing the sentiments of Justice Brandeis on sunlight as the best disinfectant. Once again, the autonomy and fairness perspectives would invite caution. Transparent to whom and for what purposes would be the questions raised under these two perspectives. The autonomy perspective would be concerned with who is allowed access to records and for what purposes. The fairness perspective would be 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. This section has highlighted some of the critical issues that will be explored in greater depth through the case studies in Section Four, particularly the case studies on access to tax records and real estate assessments and on data.gov.

B. Selling, Sharing, and Data Commercialization

Legal ownership of data provides a necessary condition for transacting in data either through sharing or market transactions, but such transactions may be further limited by legal rules. Such legal rules are designed to protect interests of transacting parties. The legal structure of transactions maps onto the rules of contract, intellectual property licensing, tort law, and other regulatory fields. But there are several specific structures worth highlighting here.

First, United States law allows for restrictions on the use of data through contract law. In *ProCD v. Zeidenberg*, the Seventh Circuit held that a seller of a database can restrict access to data through a contractual limitation. Such a contract term was found not to be preempted by federal intellectual property law and was held to complement the interests of intellectual property holders. This decision has been a controversial one since it was first handed down in 1996, but

---

125 86 F.3d 1447 (7th Cir. 1996).
with some narrow exceptions regarding contractual formation, the ruling is generally accepted.\textsuperscript{126} The rationale behind this contractual protection for 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. 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 affectsively binds any purchaser to the 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 purchasers may not in practice feel the difference between being bound by contract and being bound by copyright law. In addition, the use of arbitration 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 Four, is the extent to which technological protections, such as encryption, can be used to protect data.

However, legal rules also 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 with the Health Insurance Portability and Accountablity Act of 1996 (HIPAA) regulations providing one example of detailed government regulations protecting privacy interests in health data.\textsuperscript{127} The tort of intrusion generally protects against invasions of one’s reasonable expectation of privacy, and the related tort of false light prevents reputation harming uses of data. The right of publicity, one of the original common law torts of privacy, has taken on a life of its own and protects against the unauthorized commercial exploitation of one’s public persona. For all of these torts, the First Amendment places some limitations on tort claims against reporting or communication of private data and information. Finally, the Fourth Amendment would place limits on the government, whether acting in its crime enforcement or its regulatory capacity, from unwarranted searches and seizures and data and information obtained directly or indirectly from such searches or seizures.

With advances in genetic technologies, claims against the misuse of personal genetic information have grabbed the attention of scholars and policymakers. Statutes against genetic discrimination prevent unequal treatment in the use of genetic data.\textsuperscript{128} Collection of tissue samples and cell cultures are not governed by property law,\textsuperscript{129} 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 what is in effect personal genetic information under legal scrutiny.\textsuperscript{130} 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.


\textsuperscript{127} See Solove, supra note 5 at 69-70 (analyzing HIPAA).


\textsuperscript{129} See Washington University v. Catalona, 490 F.3d 667, 674 (8th Cir. 2007)(biological materials treated as inter vivos gifts from research participant to university).

\textsuperscript{130} Moore v. Regents of the University of California, 793 P.2d 479 (1991).
against researchers who misuse such materials. Such potential claims shape transactions in genetic data and the increased use of informed consent in hospital and medical research settings.\textsuperscript{131} The case of Henrietta Lacks recently received scholarly and media attention.\textsuperscript{132} A cell line obtained during her treatment for ovarian cancer had been used by several generations of medical researchers and students after her death. Her example 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.\textsuperscript{133}

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 contracting regime to protect data through contractual restrictions, as we will see in greater depth in the case study on open records and real estate assessment data in Section Four. The autonomy and fairness perspectives provide support for several limitations on data based transactions, especially through protections for privacy and for 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 of commercialization of data go beyond the facts of Sorrell. After a discussion of the case studies, I return to the Sorrell decision and 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 paper 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. The case study focuses on 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 state 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 under access to federal government data and transparency at the federal level. This case study shows how regulation controls 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 which provides

\textsuperscript{131} See Skloot, supra note 12 at 207-211 (describing benefit sharing agreement between hospital and patient permitting harvesting of patient’s biological materials).
\textsuperscript{132} See Skloot, supra note 12 at 305-306.
\textsuperscript{133} See Ghosh, supra note 12 at 107.
limits on 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.

1. Real Estate Data

In 2002, WireData, a private company providing information for realtors on real property that is up for sale, submitted a request under Wisconsin’s Open Records Act on various municipalities in the State of Wisconsin for tax and real property assessment data. The goal was to include these data in online real estate listings so that potential buyers would have more information for properties that might be of interest for purchase. Some municipalities informed WireData that they did not have the data which was collected by Assessment Technologies, a private company that compiled and archived the assessment data. WireData requested the data from Assessment Technologies, which 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 applied, the assessment records were exempt from the Act because they were protected by copyright law. WireData sought a declaratory judgment on the copyright status of the real estate data that it was seeking. A magistrate judge on the Seventh Circuit ruled in favor of Assessment Technologies. WireData appealed successfully to the Seventh Circuit, which ruled in an 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 the use of copyright law to deny access to unprotected data constituted copyright misuse, an equitable doctrine that limits the right of a copyright owner that 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 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 web site of real estate information. In light of the Seventh Circuit ruling, Assessment Technologies decided to comply with the request for electronic copies of the data by providing electronic files of the assessments in a .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 the 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 could not deny a request for records by referring the requestor to a private third party. Furthermore, a private third party could be subject to an open records request when hired to collect data for the public entity. However, the Wisconsin Supreme Court ruled that WireData’s request for electronic records was met by the transfer of

---

135 Id. at 743.
136 Id. at 745.
137 Assessment Technologies of WI, LLC v. WIREdata, Inc., 350 F.3d 640, 643 (7th Cir.2003)
138 WIREdata, supra note 134 at 755.
the .pdf files since they were in the requested electronic form. The high court in other words strictly construed the original request and held that it was met.139

Among the lessons from the 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 records created by state entities. The state cannot escape obligations of the Open Records Act through privatization. But the Act also imposed obligations on private citizens in framing requests and, under some situations, in bearing the burden and cost of disclosing the requested information. In this way, a market-like institution is created for the generation and dissemination of data.

When government data collection and open records obligations are understood a 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 dispute illustrates the liberal perspective on data and information access, the notion that markets and politics work together to protect and ensure individual rights. 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, the law recognizes that data protected through copyright, trade secret, or patent can be protected from request by private citizens. This private sphere within transparent government serves to protect those who process data into economically valuable products protected 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 and a proprietary sphere that supports markets. The two work together to protect individual liberty.

The 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, privacy and reputational interests may be protected by subsidiary laws, such as defamation. But such protections may not be fully adequate under an autonomy or fairness perspective. Protection of these interests would require thinking beyond individual freedom and to shape the normative framework within which market transactions and politics occur. The open source proposal in the last subsection addresses these concerns.

The first case study 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

139 Id. at 757.
is openly accessible. This case study illustrates similar problems with the liberal perspective, but also demonstrates 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 data collected by the government for ready access. This trend by governments, state and federal, to publish data has led to private entities to create search tools to identify data, aggregate it, and analyze it. Google.com, for example, provides Public Data Explorer, an application designed for journalists and researchers to analyze public databases to discover relationships and correlations in data gathered by government agencies. Within this environment, it is not surprising that statisticians are viewed as modern day rockstars, with their ability to make data speak and use mathematical techniques as investigative tools. Against this environment, however, is the WikiLeaks controversy where the values of accountability conflict with the pragmatic concerns of security and the need for government secrecy and expertise. Government data raises a host of controversies which requires consideration of the normative foundations for the governance of data.

Adding to the controversy, public data bases serve as resources for commercialization and the development of socially useful applications. A notable example is the use of publically available information on train schedules to create a smartphone application to check on arrival and departure times for connecting trains. Software engineers explore how available data on energy usage to identify areas where conservation efforts can be targeted. These examples illustrate how private enterprise can be spurred by government investment in data collection and dissemination. Government collected data are made available to the public as resources for new applications and information based products. However, unsettled issues exist 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 resulted in settlements favorable to the government.

An important initiative of the Obama Administration is the creation of the website data.gov which provides access to data collected by federal agency and not otherwise protected as a government secret for security reasons. The website terms of use illustrate both a commitment by the current administration to government transparency and protection of private interests. For example, the data policy statement which is linked from the data.gov homepage limits data access that conflicts with protections for national security and privacy as defined by

---

140 Available at http://www.google.com/publicdata/home.
the Office of Management and Budget (OMB) guidelines.\textsuperscript{143} The statement also imposes requirements for citation to datasets downloaded from the site and obligations on the collecting government agency for data retention and data integrity.\textsuperscript{144} 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.\textsuperscript{145}

Although the statement is entitled “Data Policy,” these provisions on Data.gov can be understood as terms of use which define the rights and duties of those who download from the site and make use of the data. Arguably they can be enforced by the government against individuals who violate these terms.

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.\textsuperscript{146}

The distinction between data and information is critical here for a number of reasons. Under the terms of the site, public access and agency maintenance obligations apply to data, but not to information. The distinction pertains to both what users are allowed to do and how the government must carry out its duties. 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 that “data become ‘information’.” The implication is 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, quoted in the previous paragraph. 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.

\textsuperscript{144} Id.
\textsuperscript{145} Id.
\textsuperscript{146} Id.
The distinction between data and information is a critical one for the governance of the datasets made publically available through data.gov. The distinction allows users some proprietary protection for information generated from the publically 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. But there is an interpretation consistent with the values of transparency and open government. Under the terms of Data.gov, individuals are allowed 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 has exclusionary rights. This information becomes the tools for individual users as they engage with others in deliberation and participation in open government. What I would argue is that the terms of use under Data.gov envision an autonomous individual that can engage in the public sphere but can also retreat. This argument is consistent with the privacy protections that the terms of use permit. In short, the values reflected in Data.gov are consistent with the autonomy perspective described in the previous section.

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 another rationale for the exclusion. National security is information that is the product of the government processing of data. It 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 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 in order to determine the boundaries of proprietary rights. It is not clear

148 See supra note 27.
how the public and private spheres are being drawn for the purposes of data governance under the terms of the Data.gov site.

The reason for this lack of clarity is the normative thin notion of autonomy that informs data policy. Open government initiatives like Data.gov assume the zone of autonomy is defined by data processing, a sphere of cognitive decision making through which data is sorted, analyzed, and interpreted in order to aid the investigative process in which the user is engaged. The autonomy perspective leaves open the purpose behind the investigative process. It could be an aid in deciding whom to vote for, to assess the effectiveness of government policy, to identify waste and corruption, or simply to satisfy curiosity. The autonomy perspective reflected here is neutral as to goals, but seeks to provide space for the autonomous individual to define itself through engagement with data. Open government of this sort is a discussion of facts without necessarily an engagement in deeper values or commitments. There is of course a value in the discussion of facts. While such a notion of autonomy may be thin, it may also be essential for a richer development of the self. What Data.gov demonstrates, perhaps, is that the life of data may be a thin one, even if it is one that informs much of contemporary society and economy.

While 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 I discuss later in this section. As I suggest, the terms of use from the web site 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 U.S-style Freedom of Information Act legislation. Such initiatives demonstrate that the concerns raised in this Article about commercialization of data are global and require a comparative perspective.

Two major differences 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 that treats the matter 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 protects database and data more strongly than the United States with the result that private incentives for the commercialization of data may be

151 See Reichman & Samuelson, supra note 46 at 97-98.
These two differences 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 European Database Directive, point out that one of the purposes for implementation of 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 U.S. Congress had rejected the adoption of European style database protections several times during the 1990’s. Instead, the U.S. relies on the regulatory environment described in Section Two of this Article, a mix of weak intellectual property and common law protections. The 2005 Directive 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 the issue of how the U.S. has managed to dominate globally despite lack of strong domestic protection 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 information in the gathering and collecting of data. The right extends to the prevention of 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 television programming guide 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,

---

152 See id. at 84-85 (describing threat to public domain from overprotection of data).
153 See id. at 72-73.
155 See supra note 46.
156 See supra note 154 at 23.
compelling access to data and databases raised questions about the scope of the sui generis right created by the Directive.\textsuperscript{158}

Additional uncertainty arose from the Court’s decision in the William Hill case involving access to horse racing data.\textsuperscript{159} The court ruled that an Internet database consisting of statistics pertaining to race horses was not protected by the Directive 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 was more far reaching 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 was holding that data constructed, perhaps through some statistical or interpretative process, would not be protected by the Directive, whose goal was to protect investment in the costly gathering of data. The decision was particularly confusing because of the suggestion 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 horse racing statistics) could obtain sui generis rights in the secondary database. The 2005 study concludes that the uncertainty created by this 2002 decision may have undermined the purpose of the Directive to create stable and certain property rights in databases.\textsuperscript{160}

The experience with the European Database Directive illustrates some of the problems with establishing strong property rights in data. Underlying competition law policies limit the strength of this property right. Furthermore, there are underlying and difficult questions of what constitutes data which are unresolvable. These issues also arise in the United States context, as the cases of WireData and Data.gov illustrate. The distinction that the European Court drew between constructed and unconstructed data in the William Hill case parallels the discussion of the demarcation between data and information in the context of Data.gov. If the world of data commercialization is moving towards 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 prevalent approaches to data and databases in Europe and the United States suggests that a thin protection for databases and generally open access to data 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 that allow consumers to access and manipulate information. Competition may not be the pure textbook kind of many firms and many consumers driven by price signals. The types of competition we see in database markets may be more of the oligopolistic type with few firms competing on different margins for consumers. Nonetheless, such competition can be vibrant for an industry and may be consistent with other non-economic 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 structures and their protection for values of freedom and autonomy, the third, and final, case study shows how competition informs the

\textsuperscript{158} See supra note 183 at 23.
\textsuperscript{159} See supra note 154 at 28.
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 majority and dissent operate with an antiquated notion of the marketplace of ideas. According to this antiquated notion, individuals communicate to others in order 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 the Virginia 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 uphold the regulation as reducing the influence of certain dominant speakers in the marketplace and thereby equalizing the terms of deliberation and debate. For both the majority and dissent, the marketplace of ideas, and hence the First Amendment, is about deliberation, communication, and debate.

Such a model is appropriate for certain forms and styles of communication, particularly ones involving political tracts and books. But as I have argued in this Article, the marketplace of ideas is really a marketplace of information and data. Deliberation and debate are not the engine of the contemporary marketplace. Rather, innovations such as blogs, social media, and other platforms serve as a ways of communicating facts and nuggets of information. These nuggets serve as resources that can be gleaned to learn about other people, other places, and other modes of thinking. Information and data are created and accumulated for their own sake and not to persuade others of the rightness or wrongness of particular ideas. The metaphor of the marketplace of ideas does not capture the world in which contemporary information networkers live, work, and post.

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, relevant inquiry is whether the marketplace for information and data as 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 as comprehensive but also as the 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, I do not want to be the person responsible for shutting down the Internet either or any other mechanism for collecting and disseminating information. Nothing I am advocating 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 are touched by the flow of information and data.

As one example of how the analysis of this Article need not threaten the Internet, I present a proposal for data commercialization in the next section. 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 in Sorrell is one solution, but perhaps an unrealistic one. This Article has pointed to emerging problems posed by data commercialization, some traditional such as privacy and security, others novel such as the boundaries between government management of data and private uses. The case studies 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. This proposal, implemented 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 does not address any specific problem raised by data commercialization. Instead, the goal is to provide a framework for regulation that can be used to limit some of the potential problems posed by data commercialization. The model 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. While the specific terms were the subject of criticism in the previous section for focusing on a very narrow conception of the autonomous user, the model of terms of use is an appropriate one. Furthermore, 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 has demonstrated that we live in a world where data commercialization will be standard practice. My proposal is to use licensing terms that can effectively regulate the practices of commercializing data. the ideas build on the work of Professor 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 paper.

The licensing terms can be implemented in several ways. The first is to have state and federal governments that make data available through web sites 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 in this Article should serve as a starting point for identifying and addressing issues. 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 dissemination of data generated and collected from a specific site. Competition among 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? I propose the following terms at the minimum:

- 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;
- 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 of 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 in order 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 in order to structure the regulatory response more creatively and flexibly.

V. The Future of the Marketplace of Information and Data

As the Supreme Court’s analysis in its 2011 Sorrell decision, 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 addressed adequately either the normative framework for the marketplace of ideas or the regulatory structures that make the generation and exchange of data possible. This paper fills both of those gaps.

The normative framework for the marketplace of ideas requires a consideration of class liberal principles with regards to freedom and rules, of autonomy with respect for individual notions of privacy, and of fairness with respect to distribution across individuals. The First Amendment as a check on communication of data and information needs to adequately balance these three notions with more weight applied to one or two of these principles depending upon the context. The majority in Sorrell emphasized the concept of autonomy by deciding in favor of a laizzez-faire view of the marketplace of ideas. The dissent leaned towards the classic liberal position with an emphasis on the market harms that arise from a concentrated market for data arising from the market power of pharmaceutical companies in targeting doctors. Neither adequately balanced the principles nor fully considered fairness to patients and consumers in the analysis. 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.

Neither majority nor the dissent addressed the network of government regulation that makes communication of data possible. The dissent came closer to acknowledging this network but focused largely on food and drug regulation. As set forth in this paper, however, the network 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 give 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 paper details this set of information 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.

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 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. This
paper has made the case that the dissent’s views may become more salient once we recognize the complex web of governmental interests that inform data commercialization.