
COMMERCIALIZING DATA

SHUBHA GHOSH*

I. INTRODUCTION

If the cliché of our living in an information age is true, then equally true is our lack of ability to fathom what it means to live in a world where data is the critical resource for achieving political, social, and economic values. We are told that data want to be free, and that making data freely available supports transparency, accretion of knowledge, and greater insight into personal and social behavior. At the same time, data provides a way to market products and services in more sophisticated ways that allow the supplier of goods to price discriminate close to perfectly. This article explores the conflicting legal, social, and economic issues raised by the commercialization of data.

A striking example of the problem at the heart of this article is the expansion of the 2010 Census to allow same-sex couples to list themselves as married.¹ A recent NPR report commented on how liberating this option is for the lesbian, gay, bisexual, and transgender (“LGBT”) community.² Supporters of same-sex marriage, like myself, listened to the report with hope that the ability to tick the married box on the form would provide a slim foundation for broader legal and social recognition of same-sex marriages. After all, measuring a phenomenon is one indication of recognizing it, which is a key step in understanding and legitimizing it. The same report, however, interviewed advertisers and marketing experts who welcomed the change because it allowed

* Professor of Law, University of Wisconsin Law School

¹ U.S. CENSUS BUREAU, U.S. DEP’T OF COMMERCE, FORM D-3400, 2010 CENSUS FACT SHEET FOR LESBIAN, GAY, BISEXUAL AND TRANSGENDER PERSONS 2 (2010) (“LGBT people living with a spouse or partner can identify their relationship by checking either the ‘husband or wife’ or ‘unmarried partner’ box.”), *available at* http://2010.census.gov/partners/pdf/factSheet_General_LGBT.pdf.

² *Tell Me More: 2010 Census Will Count Same-Sex Couples*, NAT’L PUB. RADIO (Nov. 25, 2009), <http://www.npr.org/templates/story/story.php?storyId=120816467>.

companies to fine tune advertising and identify new markets in which to enter and sell.³ On one level, political and social recognition necessarily implies economic recognition. Consequently, the cheering of advertisers could be a source of celebration. However, eager listeners tuning in closely may fear that economic recognition may not lead to political and social acceptance. What, then, are we to make of the new data points that the 2010 Census will uncover?

The tension the Census example illustrates reveals an important facet of the generation of data. No one can predict where new data will lead. It can be the source of good or the source of misery, depending on how the data is used. Quite patently, the Census example illustrates the fact-value distinction. Data have no inherent value and can be used in unpredictable and uncontrollable ways. This is true for any input, whether labor, machines, or natural resources. But unlike these resources, democratic civil societies operate with the assumption that data generation is in itself a valuable activity. The claim that sunlight is the best disinfectant assumes that the revealed data will prevent chicanery and corruption.⁴ Free and open inquiry, whether in the realm of the humanities, social sciences, or natural sciences, is at the center of what makes a society modern. Data are the currency for the marketplace of ideas, heralded as a constitutional value that informs democratic deliberation and copyright law, with its twin roles in encouraging creativity and commerce. Yet, are the data in fact free to be molded? Or is there a normative structure to how data is produced, used, and commercialized?

A look at www.data.gov⁵ provides some initial clues to answering these questions. The website is a forum for access to data from various government agencies.⁶ The Data Policy statement included on the site illustrates the normative foundations for access to data.⁷ Not surprisingly, national security, privacy, and regulatory concerns place limits on which data might be available.⁸ Furthermore, the policy states that distributors of data obtained from the site cannot place any restrictions

³ See *U.S. Census Recognizes Gay Marriage*, A LOS ANGELES LOVE, (Sept. 4, 2011, 7:17 PM), <http://www.losangeleslove.com/2010/05/us-census-recognizes-gay-marriage.html>.

⁴ See LOUIS D. BRANDEIS, *OTHER PEOPLE'S MONEY: AND HOW THE BANKERS USE IT* 92 (1914).

⁵ DATA.GOV, <http://www.data.gov> (last visited Aug. 9, 2011).

⁶ *About*, DATA.GOV (Aug. 19, 2011, 3:21 PM), <http://www.data.gov/about>.

⁷ *Data Policy*, DATA.GOV (Aug. 9, 2011, 2:06 PM), <http://www.data.gov/datapolicy>.

⁸ *Id.*

on downstream uses of the data.⁹ Most notably, there are no prohibitions against commercializing or selling the data to third parties.¹⁰ Effectively, the Data Policy creates an open source license for the dissemination of data. While this implicit license does not prohibit commercialization, the limitations in the license, and the ease with which data can be accessed, seemingly make commercialization opportunities unprofitable. Nonetheless, as the Census example shows, commercial possibilities always can be found. The policy statement's silence may reflect an implicit license to commercialize data generated from the site.

What is particularly telling about the website is the definition of data. To quote: "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."¹¹ This elusive definition reflects the challenges in identifying the key ingredients for commercial activity in the information age. The distinction between data and information appeals to the distinction between raw data and processed (or cooked) data, which may be an essential difference in how we define legal rights over data. Furthermore, this distinction belies another concept that is critical to the information age: the record. All states and the federal government have statutes that allow access to records, commonly known as Open Records Acts, which are one particular form in which data and information may be packaged while also imposing limits on their use.¹² Much like data itself, ways of talking about data can propagate quickly and the distinctions can often become elusive, while at the same time they can be critical in understanding the legal, social and economic control of data.

The approach in this article is critical and foundational. The goal is to identify and analyze the key variables in the commercialization of data with the ultimate aim of understanding the normative structure

⁹ *Id.* ("Data accessed through Data.gov do not, and should not, include controls over its end use.").

¹⁰ *See id.*

¹¹ *FAQ*, DATA.GOV (Aug. 9, 2011, 2:17 PM), <http://www.data.gov/faq> (quoting FED. ENTER. ARCHITECTURE PROGRAM, OFFICE OF MGMT. & BUDGET, THE DATA REFERENCE MODEL: VERSION 2.0 84 (2005), available at http://www.whitehouse.gov/sites/default/files/omb/assets/egov_docs/DRM_2_0_Final.pdf).

¹² *See, e.g.*, 5 U.S.C. § 552 (2009); N.C. GEN. STAT. §§ 132-1, 132-6, 132-6.1, 132-6.2 (2005).

for regulating the uses of data. The structure of this article is as follows: a review of some of the confounding conceptual issues raised by data leads to an analysis of how rights over data can be defined. The focus will turn to an important, and unexamined, case involving the commercialization of data, *Assessment Technologies of Wisconsin v. WireData, Inc.*, a dispute which generated opinions from the Seventh Circuit and the Wisconsin Supreme Court.¹³ The article concludes with lessons to be learned from the theoretical and case analyses.

II. DATA ENTREPRENEURS AND THE MODEL OF SHERLOCK HOLMES

Over a decade ago, Professor James Boyle identified and addressed the challenges of the legal regulation of information.¹⁴ Drawing in part on the work of economists Sanford Grossman and Joseph Stiglitz, Professor Boyle pointed out that information serves many conflicting roles in neoclassical models of the marketplace.¹⁵ On the one hand, for markets to function, information must be perfect.¹⁶ On the other hand, if markets were functioning efficiently, rents from having superior information would be arbitrated away.¹⁷ Hence the paradox: how does information become generated and distributed if returns from investing in information vanish? Framed another way, the paradox becomes: information needs to be open for markets to thrive, but to profit from information it must also be proprietary.

One response is to recognize that information is not uniform and different types of regimes may protect different types of information. For example, in a market context, information about price and product quality might be viewed as open and transparent while information about a firm's technology or a consumer's preferences may be proprietary. The scholarly literature on the legal regulation of information has moved, in part, to recognizing the granularity and diversity of information. Accompanying this move is recognition of the rich institutional environment in which information is produced and consumed. Markets for information, like all markets, exist in a rich ecosystem, and one must understand the dynamics of these institutions to appreciate the manner in which information is produced and disseminated. Joel

¹³ *Assessment Tech. of Wis. v. WireData, Inc.*, 350 F.3d 640 (7th Cir. 2003); *WireData, Inc. v. Vill. of Sussex*, 751 N.W.2d 736 (Wis. 2008).

¹⁴ JAMES BOYLE, *SHAMANS, SOFTWARE, AND SPLEENS: LAW AND THE CONSTRUCTION OF THE INFORMATION SOCIETY* (1996).

¹⁵ *Id.* at 39-40, 222-23 n.16.

¹⁶ *Id.*

¹⁷ *Id.*

Reidenberg's work on *lex informatica*, by analogy to *lex mercatoria*, is one example of this deeper understanding.¹⁸ Under *lex informatica*, technology makers and keepers determine social norms that become a model for legal regulation of information.¹⁹ Reidenberg's work views technology as the basis by which regulation occurs.²⁰

This article also examines a *lex informatica*, but does not rest its formation solely on technologists. Like *lex mercatoria*, *lex informatica* develops from market and social contexts in which rules and conventions of various degrees of formality generate governing laws.²¹ In the case of *lex mercatoria*, these governing laws regulate commercial transactions. In the case of *lex informatica*, the governing laws modulate the manner in which information is generated, disseminated, and commercialized. My premise, however, is that this is not just a matter for the technologists. As the Census 2010 and www.data.gov examples show, this is a matter of technology embedded in broader social, political, and economic contexts.²² This section draws a sketch of these contexts.

The model I have in mind for *lex informatica* is not the contemporary technology manager but Sherlock Holmes. As a detective, Mr. Holmes was a processor of data and a generator of information who yielded answers to seemingly intractable questions.²³ I will not provide any spoilers here about his sleuthing, but most readers are perhaps aware of how the detective was able to discern whole life histories from simple facts like a scar or the way in which a person entered a room. Dr. Joe Bell, the professor of Holmes' creator Arthur Conan Doyle, was legendary in being able to make diagnoses from the minutest of details.²⁴ Holmes is a data producer and a data consumer. His detection involved gathering and creating data as inputs into his deductive method and also acting as a passive consumer collecting information about the world. As a commercializer of data, Holmes made a living, of sorts, through the sale of his services. There are records of monographs he published on cigar ashes and other marginalia.²⁵ Nev-

¹⁸ Joel R. Reidenberg, *Lex Informatica: The Formulation of Information Policy Rules Through Technology*, 76 TEX. L. REV. 553, 553-55 (1998).

¹⁹ *Id.* at 571.

²⁰ *Id.* at 568.

²¹ *Id.* at 554-55.

²² See U.S. CENSUS BUREAU, *supra* note 1; DATA.GOV, *supra* note 5.

²³ ELY M. LIEBOW, DR. JOE BELL: MODEL FOR SHERLOCK HOLMES 133-34 (1982).

²⁴ *Id.* at 131.

²⁵ *Id.* at 177.

ertheless, his primary *métier* was using his craft at parsing data to solve puzzles and capture criminals.

The contemporary model of the commercializer of data starts with Sherlock Holmes. However, contemporary data entrepreneurs go beyond the mere sale of services (although there is a rich market for data forensics). Data entrepreneurs package data through websites and other media so that users of all sorts can generate more data in a seemingly exponential growth. The result is a counter-Malthusian economy where the geometric growth of population is matched by the exponential growth of data and information.²⁶ The post-modern literary genre of science fiction captures what would be Sherlock Holmes's contemporary environment. The so-called literature of information, including such prominent novelists as Thomas Pynchon and William Gass, but also popular figures like Philip K. Dick, reveal the complexities of this counter-Malthusian economy. "Information overload" is the popular buzzword, but, at a more sophisticated level, the question is how the generation and consumption of data affects the mindset. Sherlock Holmes was a methodical figure, a marriage of Cartesian rationality with British empiricism. In the counter-Malthusian economy of data, method and madness run in tandem as "supercrunchers," generating the next generation of results. Is everything connected or is all data the result of random processes? This is the question posed by novelists like Thomas Pynchon.²⁷ The lived experience is one where anything goes; all data is there to be used, processed, regressed, and parsed into a mix of innuendo, rumor, hypotheses rejected, statistical significance gauged and reported.

A picture of the counter-Malthusian economy emerges from asking questions about production, consumption, and ownership. At the outset, the line between production and consumption with respect to information is a blurry and perhaps non-existent one, as many scholars have demonstrated.²⁸ If data is just an input to produce information or more data, then a consumer of data is arguably just another producer in a never-ending value chain. Nonetheless, I use the labels producer

²⁶ See THOMAS ROBERT MALTHUS, AN ESSAY ON THE PRINCIPLE OF POPULATION, AS IT AFFECTS THE FUTURE IMPROVEMENT OF SOCIETY, WITH REMARKS ON THE SPECULATIONS OF MR. GODWIN, M. CONDORCET, AND OTHER WRITERS (1798), *reprinted in* FIRST ESSAY ON POPULATION, at 13-14 (Macmillan & Co. Ltd., 1966) (1926).

²⁷ See THOMAS MOORE, THE STYLE OF CONNECTEDNESS: GRAVITY'S RAINBOW AND THOMAS PYNCHON (1987).

²⁸ See Joseph P. Liu, *Copyright Law's Theory of the Consumer*, 44 B.C. L. REV. 397, 400-01 (2003).

and consumer here to distinguish between the questions, “Where do data come from?” and “Where do data go?”. These questions are useful in providing some order in understanding the process of commercializing data.

1. *Production.* Data are generated like manna from heaven. Sherlock Holmes is the canonical case. Anything, from the smallest fiber of hair to an innocent movement of the eyes, can be a source of data.²⁹ Data come in many varieties and forms. As the www.data.gov site illustrates, processing data can generate another level called information, which captures the inferences drawn as data are moved through a scheme of review and analysis.³⁰ Furthermore, data and information drawn from data may be fixed in various ways that can generally be referred to as records, which serve as storage and packaging of data and information. In addition, there are various ways to arrange data and information, such as through writings and publications that report on the analyst’s findings or in databases which serve to package data in ways that can be more readily parsed and analyzed for the generation of more information. Data production can best be divided into raw data and cooked data.³¹ The line between raw data and cooked data is one determined by normative decisions about how data should be regulated, as discussed in greater detail under ownership, below. However, framing the question in terms of raw versus cooked aids in understanding how normative questions of ownership and regulation are determined.

2. *Consumption.* As pointed out before, consumers of data are often intermediaries who are in turn producers of the next generation of data and information.³² In some instances, however, consumers of data may be pursuing data for their own sake. A consumer may simply want to know the answer to a question, such as one’s ancestry or the presence of a disease in one’s body, or something more mundane like the hours for a restaurant or museum. In terms of commercializing

²⁹ See LIEBOW, *supra* note 23, at 133-34.

³⁰ See FAQ, DATA.GOV, *supra* note 11.

³¹ See ROBERT W. GOMULKIEWICZ, XUAN-THAO NGUYEN & DANIELLE CONWAY-JONES, LICENSING INTELLECTUAL PROPERTY: LAW AND APPLICATION 418-19 (2008) (citing *Feist Publ’ns, Inc. v. Rural Tel. Serv. Co.*, 499 U.S. 340 (1991)) (distinguishing between “individual facts” and “original arrangements of facts”). See also *id.* at 373 (“Collecting, analyzing, and classifying enormous volumes of information and raw data equate to huge profits for those businesses able to harness its power through information and data processing.”).

³² See Liu, *supra* note 28, at 400-01.

data, consumers may often be the construct of the marketplace as data entrepreneurs define a niche within which the activity of data retrieval and usage occurs. Bulletin boards, blogs, and discussion groups are all examples of how the data entrepreneur can construct consumption in order to generate another level of data, which also creates value. As www.data.gov illustrates, simply providing access to data may create consumption, although with the case of government data, such demand may arise from political and economic needs to understand one's rights or how one's tax dollars are being spent.³³

3. *Ownership.* Data commercialization rests on identifiable legal rights and duties among parties in data-related transactions. The challenge is identifying the structure of these rights and duties. In the Sherlock Holmes example, the entrepreneurial detective pursued a business model based on the sale of his detection services partly because securing property rights in the inputs of his business was impossible. The clues on which he built were in the public, in plain sight; his craft was in identifying and analyzing them. Contemporary data entrepreneurs need to negotiate a panoply of legal regimes in order to negotiate data-related transactions. Privacy and security restrictions may protect personal and other sensitive types of data.³⁴ Trade secret law protects commercially valuable information within a firm.³⁵ While data is unprotected under copyright law, the law can protect the original arrangement, selection, and coordination of databases.³⁶ Furthermore, contract law might offer some protection to data on a bilateral basis.³⁷ Finally, patent law might offer some protection to the methods of how data are processed and accessed, although process patents are currently under scrutiny and may be subject to limitation.³⁸ In short, the ownership structure of data is a complex one. Proprietary in some dimensions and legally unfettered in others, data ownership reflects the contradictory roles of information in transactions as both the basis for open and liberal markets and an investment that is expected to generate a return.

³³ See DATA.GOV, *supra* note 5.

³⁴ See 2 ROGER M. MILGRIM & ERIC E. BENSON, MILGRIM ON TRADE SECRETS § 9.05[3][a] (2011).

³⁵ 1 MILGRIM & BENSON § 1.01; 1 MELVIN F. JAGER, TRADE SECRETS LAW §§ 3.2, 3.34 (2009).

³⁶ 2 MILGRIM & BENSON § 9.03[4][a][i][A]; 2 JAGER §§ 10.6, 10.7.

³⁷ See 1 MILGRIM & BENSON § 4.02[1][a].

³⁸ See 2 MILGRIM & BENSON § 9.02[4][a]; 2 JAGER, *supra* § 10.4.

With this theoretical analysis of the counter-Malthusian world of data and information, we turn next to a case example to illustrate the complexities of commercializing data.

III. TAX ASSESSORS VERSUS REAL ESTATE AGENTS

The case of *Assessment Technologies of Wisconsin v. WireData, Inc.* illustrates the interweaving set of laws that govern the commercialization of data and highlights the interaction between government data collection, disclosure obligations, and profit-driven commercial activity.³⁹

WireData is a Wisconsin company owned by Multiple Listing Services, Inc. and engaged in the advertising of properties whose owners are represented by licensed realtors in the state.⁴⁰ Part of its business rests on providing comparable information on the properties for sale so that sellers can effectively communicate the valuable characteristics of their properties and buyers can assess and compare properties based on price and quality.⁴¹ In 2001, WireData made a request under Wisconsin's Open Records Act for tax assessment data over a multi-year period collected and held by municipalities.⁴² Although the company did not state its motivation for requesting these public records, it is clear from its business model that WireData wanted to include this information on its website for real estate listings.⁴³ The municipalities did not immediately honor the request on the ground that a private entity called Assessment Technologies, whose agents assessed properties based on inspections and collected data on the characteristics of assessed lots in an electronic database, collected and maintained the data.⁴⁴ The municipalities referred WireData to Assessment Technologies as the proper party to whom they should make their Open Records request and argued that provisions of the Open Records Act that excluded trade secrets and copyrights from disclosure exempted the real estate assessments from the Open Records Act.⁴⁵ When WireData made its request for the electronic records to Assessment Technologies, Assessment Technologies brought suit in district court

³⁹ *Assessment Tech. of Wis. v. WireData, Inc.*, 350 F.3d 640 (7th Cir. 2003); *WireData, Inc. v. Vill. of Sussex*, 751 N.W.2d 736 (Wis. 2008).

⁴⁰ *Assessment Tech.*, 350 F.3d at 642; *WireData*, 751 N.W.2d at 741.

⁴¹ *Assessment Tech.*, 350 F.3d at 642.

⁴² *Id.*; *WireData*, 751 N.W.2d at 741, 746, 747.

⁴³ *See Assessment Tech.*, 350 F.3d at 642; *WireData*, 751 N.W.2d at 741.

⁴⁴ *Assessment Tech.*, 350 F.3d at 642; *WireData*, 751 N.W.2d at 742, 746, 748.

⁴⁵ *Assessment Tech.*, 350 F.3d at 641-42; *WireData*, 751 N.W.2d at 742, 744-45, 746, 748.

claiming that the records were protected as trade secrets and under copyright.⁴⁶

The District Court, in an order issued by the magistrate, found for Assessment Technologies, and WireData appealed this ruling to the Seventh Circuit, which reversed.⁴⁷ In a terse and tightly reasoned opinion authored by Judge Posner, the Court of Appeals stated in dicta that the trade secret claim was incomprehensible since Assessment Technologies' website disclosed both the data and the data formats that Assessment Technologies attempted to protect, precluding their protection as trade secrets.⁴⁸ In addition, the Court of Appeals rejected protection of the records under copyright law.⁴⁹ Assessment Technologies' argument rested on protection of its database of real estate assessments under copyright law.⁵⁰ The electronic transfer of the records, as WireData had requested, could entail a transfer of their protected database, particularly their choices of how to arrange, select, and coordinate the data.⁵¹ Consequently, the request for the database itself was exempted from the Open Records Act.⁵²

Judge Posner's reasoning rested on the well-accepted proposition that copyright law does not protect facts or data but only the way in which facts and data are expressed.⁵³ Assessment Technologies' argument was effectively extending copyright protection from the database to the unprotected data. The Court of Appeals ruled that this use of copyright law could constitute copyright misuse as an inequitable extension of the copyright grant to subject matter unprotected by copyright law.⁵⁴ The Court's reasoning rested on copyright policy, and its ruling that Assessment Technologies had to respond to WireData's request implicitly placed a limitation on the protection under the Open Records Act for copyrighted material.⁵⁵

Pursuant to the Seventh Circuit's ruling, Assessment Technologies transferred the requested electronic records to WireData in the form

⁴⁶ *Assessment Tech.*, 350 F.3d at 642; *WireData*, 751 N.W.2d at 744-45.

⁴⁷ *Assessment Tech.*, 350 F.3d at 642, 648.

⁴⁸ *Id.* at 642.

⁴⁹ *Id.* at 644-45.

⁵⁰ *Id.* at 642.

⁵¹ *Id.* at 643.

⁵² *Id.*

⁵³ *Id.*

⁵⁴ *Id.* at 646-47.

⁵⁵ *Id.* at 643-45.

of electronic PDF files, readable by Adobe Acrobat.⁵⁶ Such forms did not allow WireData to edit or manipulate the underlying data, but did allow them to see the collected information and in theory reproduce the bytes of information from the PDF files through either mechanical inputting or electronic scanning. WireData's view was that this response by Assessment Technologies failed to comply with the Open Records Act.⁵⁷ Consequently, WireData brought suit in Wisconsin state court against the municipalities and Assessment Technologies for violations of the Wisconsin Open Records Act.⁵⁸ The company had claims based on the municipalities referring WireData to serve its request on Assessment Technologies and based on the electronically inaccessible format of the PDF files in which the request was satisfied.⁵⁹ The Wisconsin Supreme Court found in favor of Assessment Technologies on the format issue.⁶⁰ In effect, WireData had asked for the electronic form of the data and had received it in an electronic format.

The dispute over tax assessment records highlights several issues raised by the commercialization of data. First is the mix of private and public entities in the collection and dissemination of data. As is quite common, the state entity has outsourced data generation and analysis to a private entity, which relies on intellectual property law to protect its creations from misappropriation and unauthorized use.⁶¹ In contracting with private companies like Assessment Technologies, however, the government cannot abdicate its obligations under the law to the public.⁶² In the WireData dispute, the Wisconsin Supreme Court made it clear that municipalities cannot abdicate their duty under the Open Records Act, and presumably under other statutory obligations, by referring private citizens to the private firm with whom the state is contracting.⁶³ More importantly, the public obligations fall on private entities when they are engaged in the collection of data which the public has a right to access.⁶⁴

⁵⁶ *WireData*, 751 N.W.2d 736, 758-59.

⁵⁷ *Id.* at 759.

⁵⁸ *Id.* at 739.

⁵⁹ *Id.*

⁶⁰ *Id.* at 759.

⁶¹ *Assessment Tech.*, 350 F.3d at 642-43; *WireData*, 751 N.W.2d at 741.

⁶² *Assessment Tech.*, 350 F.3d at 647; *WireData*, 751 N.W.2d at 762.

⁶³ *WireData*, 751 N.W.2d at 757.

⁶⁴ *See id.* at 762 (“[A] municipality may not avoid liability under the open records law by contracting with an independent contractor.”). *But see id.* (“[T]he independent contractor assessor is not a proper recipient of an open records request.”).

Second, formalism in how data are packaged and made available may burden access to data. One lesson from the WireData case is that the request for data must be specific because the state can meet the request on its literal terms. WireData ends up being hoisted on its own petard because its request for electronic forms of data was not specific enough to include electronic forms that are readily manipulable.⁶⁵ Whether the Open Records Act guarantees data access in a specific form is left to be seen. Reading the Wisconsin Supreme Court's opinion broadly, one might conclude that the state and its agents must satisfy the request for data on all its terms.⁶⁶ However, in most jurisdictions, an Open Records Act does not create an absolute right to access, but a right based on reasonable terms.⁶⁷ Across the states, courts have read a reasonableness element into the right to access, where reasonableness includes costs of replication and collection.⁶⁸ Whether reasonableness includes other factors, such as the protection of reasonable investments in the collection of data, is yet to be determined. Companies like Assessment Technologies seek some protection for the investment made in creating their databases. Their resistance to turning over data represents a concern for turning over their "crown jewels," namely the data structures and algorithms used to organize and analyze the collected data. The Wisconsin Supreme Court's opinion may not represent deference to the terms on which the requester frames the demand for data, but rather a shift of the burden to the requester to present a reasonable request for the valuable data.

Third, the case represents the murky property rights that surround data. Copyright law cannot protect data, but data can be protected as a trade secret.⁶⁹ In the WireData dispute, the Seventh Circuit was skeptical about trade secret protection because Assessment Technologies had not taken the steps needed to protect the data as a trade

⁶⁵ *Id.* at 759.

⁶⁶ *See id.* at 760 (reasoning that where WireData made a proper records request, the issue becomes whether municipalities fulfilled the request).

⁶⁷ *See, e.g.*, 5 U.S.C. § 552(a)(3)(A-C) (2009); N.C. GEN. STAT. § 132-6(a) (2005).

⁶⁸ *See, e.g.*, *State Emps. Ass'n of N.C. v. N.C. Dep't of State Treasurer*, 695 S.E.2d 91, 95 (N.C. 2010); *Gannett Pac. Corp. v. N.C. SBI*, 595 S.E.2d 162, 164 (N.C. Ct. App. 2004) ("Therefore, it is the policy of this State that the people may obtain copies of their public records and public information free or at minimal cost unless otherwise specifically provided by law." (quoting N.C. GEN. STAT. § 132-1(b) (1995))).

⁶⁹ *See generally* 1 MILGRIM & BENSON, *supra* note 34, § 1.01; 1 JAGER, *supra* note 35, §§ 3.2, 3.34.

secret.⁷⁰ However, the court's reaction does not rule out trade secret protection through the steps required by state law. Furthermore, even if data are not protected, databases are protected.⁷¹ This distinction is well-worn in the legal scholarly literature. Legal scholars and policy makers in the United States have expressed skepticism about a European style protection for databases precisely because of the concern of data being locked up and made inaccessible.⁷² However, practitioners often grimace at the seemingly facile legal distinction between data and databases. The term "record" is one example of this confusion. Is an individual record a piece of data, unprotected by copyright, or a database that is subject to protection? Put another way: is a piece of data part of a record or a record itself?

Neither the Seventh Circuit nor the Wisconsin Supreme Court defined any of these terms precisely. The Seventh Circuit avoided addressing the question of what constitutes a record and relied on a commonsensical distinction between data and database.⁷³ In the WireData dispute, the individual characteristics of the plot of land constituted data, and it was clear that the company wanted these individual characteristics through their request.⁷⁴ Framed another way, these individual characteristics were the elements that constituted Assessment Technologies' database and therefore the line between data and database could be resolved in this case. In other cases, however, courts have ruled that index numbers, such as pricing information, can constitute a protected database consisting of a composite of data.⁷⁵ So, for example, the blue book value of a car or a stock index, even though seemingly a piece of data, represents a composite of more primitive information that is aggregated into the final number. When speaking of data, what is unprotected input and what is protected output may be an ad hoc determination.

Finally, the WireData dispute raises the question of the extent to which data is a public resource or infrastructure, to use Professor

⁷⁰ *Assessment Tech. of Wis. v. WireData, Inc.*, 350 F.3d 640, 642 (7th Cir. 2003).

⁷¹ See 2 JAGER, *supra* note 35, § 9.3 ("In fact, trade secrets are by far the most commonly used method for protecting software." (citation omitted)). See also *id.* § 9.4; 1 MILGRIM & BENSON, *supra* note 34, § 1.09, -[5][b] n.325.

⁷² See LAWRENCE LESSIG, *FREE CULTURE: HOW BIG MEDIA USES TECHNOLOGY AND THE LAW TO LOCK DOWN CULTURE AND CONTROL CREATIVITY* 81-82, 146 (2004).

⁷³ See *Assessment Tech.*, 350 F.3d at 644.

⁷⁴ *Id.* at 642, 644.

⁷⁵ See *Bd. of Trade of Chicago v. Dow Jones & Co.*, 456 N.E.2d 84, 90 (Ill. 1983).

Frischmann's phrase,⁷⁶ and the extent to which data is a private resource that individual users must collect. WireData's use of the available tax assessment records avoided a costly and wasteful duplication of effort in data collection and generation. At the same time, Assessment Technologies had an investment and business model that it wanted to protect through reasonable means. As we have seen with public data, like the Census or information available on www.data.gov,⁷⁷ motivation for commercial gain does not limit access to data. Although terms of use limit the ability to lock up data, they do not prohibit commercialization. By permitting commercialization, access to public data effectively subsidizes commercial enterprises. This implicit subsidy creates a lack of parity between public-private enterprises and purely private ones. For example, WireData can protect its databases through intellectual property law and its data through trade secret law. There is no equivalent, however, to the Open Records Act that would allow a company to access data or databases from WireData.

This lack of parity, which may or may not be the right result, reflects a normative determination both about what types of data should be public and about the mix of private and public enterprises in commercializing and disseminating data. The difficult issue is how to make this normative determination. In the famous *International News Service v. Associated Press* case, the Supreme Court's recognition of a quasi-property right in hot news reflected an ideological choice but also a determination that companies should be encouraged to investigate and discover the news, rather than simply wait for a competitor to report it.⁷⁸ This latter determination reflects the need for competition in news gathering. More contemporary approaches to the problem of hot news rely, with greater justification and less ideological sway, on first-mover advantage in order to create incentives for news gathering and reporting.⁷⁹ First-mover advantage has many benefits, including monitoring costs. The difficult question is how to develop a more complete and practical approach to creating normative justifications for the regulation of data. The WireData dispute offers a salient backdrop to raise and investigate these questions.

⁷⁶ Brett M. Frischmann, *An Economic Theory of Infrastructure and Commons Management*, 89 MINN. L. REV. 917, 918-19 (2005).

⁷⁷ See U.S. CENSUS BUREAU, *supra* note 1; DATA.GOV, *supra* note 5.

⁷⁸ *Int'l News Serv. v. Assoc. Press*, 248 U.S. 215, 236 (1918).

⁷⁹ See, e.g., *Barclays Capital, Inc. v. Theflyonthewall.com, Inc.*, No. 10-1372, 2011 WL 2437554, at *18 (2nd Cir. 2011); *Confold Pac., Inc. v. Polaris Indus., Inc.*, 433 F.3d 952, 960 (7th Cir. 2006); *NBA v. Motorola, Inc.*, 105 F.3d 841, 852-53 (2nd Cir. 1997).

IV. RAW DATA AND COOKED DATA AS NORMATIVE TROPES

This article investigates the challenges in commercializing data. As opposed to previous scholarly investigations of this challenge, my approach is not to fall back solely on technological solutions through *lex informatica* or on common law theory of property and tort. Instead, this article investigates the commercialization of data through actual cases in which access to data squarely raises the tensions between democratic accountability and market autonomy. I will now examine a few additional points about what we know from this investigation.

First, property rights in data undercut important values of access and may even be inconsistent with markets for data. At the same time, no limits on access through some defined property rights may in effect create subsidies for some enterprises, particularly ones that do not rely on public-private ventures. The difficulty is understanding the normative foundations for why we want access and why we want limitations. Those foundations may depend upon the type of data and the type of uses.

Second, understanding these normative foundations for data access requires accepting the commercialization of data. Imposing bans or restrictions on data commercialization may not be realistic and may not be desirable. Although personal privacy is an important trump in the access and dissemination of data, data can be refined and processed in different ways in order to obtain a commercial reward. As the example of the 2010 Census that started this article demonstrates, this nod to data commercialization can be troubling.⁸⁰ If one goal in a pluralistic democracy is to respect and protect the autonomy of all groups, then one has to recognize that personal autonomy has civil, political, personal, as well as economic dimensions. General prohibitions on targeted advertising that is harassing or otherwise harmful can be justified. Nevertheless, if advertisers use the 2010 Census information on same-sex marriage to market towards same-sex couples and LGBT individuals,⁸¹ then one can hope that this is a first step to recognizing broader civil and personal rights. Therefore, while the NPR report⁸² initially troubled me, banning the commercialization of data is not a measured response.

⁸⁰ See U.S. CENSUS BUREAU, *supra* note 1; *Tell Me More*, *supra* note 2.

⁸¹ See U.S. Census Recognizes Gay Marriage, *supra* note 3.

⁸² See *id.*

Finally, as this last example shows, the topic of commercializing data raises critical questions about the relationship between markets and democracy. Data may be free in a political sense, but not so in an economic sense. What notion of freedom should triumph? This tension is what thinkers like Lawrence Lessig capture in the trope of free beer versus free speech.⁸³ It can also be understood through the tropes of raw data and cooked data. If data in their purest and unprocessed form need to be accessible for political deliberation and economic transformation, then how cooked do data have to be before limits on freedom become acceptable? This project is an attempt to answer that deeper question by looking at how data has been and continues to be commercialized.

⁸³ See LESSIG, *supra* note 72, at 2-3 (“[W]e come from a tradition of ‘free culture’ – not ‘free’ as in ‘free beer’. . .but ‘free’ as in ‘free speech,’ ‘free markets,’ ‘free trade,’ ‘free enterprise,’ ‘free will,’ and ‘free elections.’”).