A FREE RIDE: DATA BROKERS' RENT-SEEKING BEHAVIOR AND THE FUTURE OF DATA INEQUALITY
(forthcoming Spring 2018 publication with Vanderbilt Journal of Entertainment & Technology Law)

Laura Palk[1] and Krishnamurty Muralidhar[2]

Abstract

Historically, researchers obtained data from their independent studies and government data. However, as the public outcry for privacy regarding the government's maintenance of data has increased, the discretionary release of government data has decreased or become so anonymized that its relevance is limited. Research necessarily requires access to complete and accurate data. As such, researchers are turning to data brokers for the same, and often more, data than they can obtain from the government. Data brokers base their products and services on data gathered from a variety of free public sources, and via the governmentally-created Internet. Data brokers then re-categorize the existing, free data, and combine it with privately collected data. They sell the linked data at a profit while simultaneously preventing the public, whose data they sold, from learning how the data was gathered based on their trade secret protections. To our knowledge, research has not explored data brokers' rent-seeking behavior and how it will further inequality in accessing credible data—or "data inequality." We contend data brokers' sale of personal data, without a concomitant federal mission to ensure cost-free access to such data for research and public access purposes, potentially will lead to biased or inaccurate research results, furthering the interests of the educated wealthy at the expense of the general public. To resolve this growing data inequality, we recommend a variety of legal and voluntary solutions.

I. Introduction

"'Rent seeking' is one of the most important insights in the last fifty years of economics and, unfortunately, one of the most inappropriately labeled. Gordon Tullock originated the idea in 1967, and Anne Krueger introduced the label in 1974. The idea is simple but powerful. People are said to seek rents when they try to obtain benefits for themselves through the political arena. They typically do so by getting a subsidy for a good they produce or for being in a particular class of people, by getting a tariff on a good they produce, or by getting a special regulation that hampers their competitors. Elderly people, for example, often seek higher Social Security payments; steel producers often seek restrictions on imports of steel; and licensed electricians and doctors often lobby to keep regulations in place that restrict competition from unlicensed electricians or doctors."[3]

Leading economists agree that rent-seeking is detrimental to a free market economy and leads to a decline in growth.[4] Rent-seeking behavior creates more income inequality, and thus, other forms of inequality by bending the "rules of some system to shuttle more compensation [to the

[1] Lecturer, Legal Studies and Accreditation and Assurance of Learning Coordinator, University of Oklahoma Price College of Business and Assistant Adjunct Professor, University of Oklahoma College of Law.
[2] Professor, Marketing & Supply Chain Management, University of Oklahoma Price College of Business.
[3] *See* ***Rent Seeking,*** THE CONCISE ENCYCLOPEDIA OF ECONOMICS LIBRARY OF ECONOMICS AND LIBERTY *available at* http://www.econlib.org/library/Enc/RentSeeking.html (last visited April 3, 2017).
[4] *See* Jim Tankersley, *A big shot venture capitalist says we need inequality, What do economists say?* WASHINGTON POST (Jan. 14, 2016), *available at* https://www.washingtonpost.com/news/wonk/wp/2016/01/14/what-silicon-valley-doesnt-understand-about-inequality/?utm_term=.8f8413012607.

wealthy]."[5] Generally speaking, if a business is not adding value to the economy and reaping financial rewards anyway, it is rent-seeking.[6] To our knowledge, research has not explored data brokers' rent-seeking behavior and how it will further inequality in accessing credible data—or "data inequality."[7] We contend data brokers' sale of personal data, without a concomitant federal mission to ensure cost-free access to such data for research and public access purposes, potentially will lead to biased and inaccurate, or at least uncorroborated and unchallenged, research results, furthering the interests of the educated wealthy at the expense of the general public.

Historically, researchers obtained data from their independent studies and government data.[8] However, as the public outcry for privacy regarding the government's maintenance of data has increased, the discretionary release of government data has decreased or become so anonymized that its relevance may become limited.[9] Research necessarily requires access to complete and accurate data.[10] As such, researchers are turning to data brokers for the same, and often more, data than they can obtain from the government.[11] Data brokers base their products and services on data gathered from a variety of free public sources, and via the governmentally-created Internet.[12] Data brokers then re-categorize the existing, free data, and combine it with privately collected data.[13] They sell the linked data at a profit while simultaneously preventing the public,

---

[5] *Id.*

[6] *Id.* As part of income inequality's negative effect on the economy, economists have concluded the rich are enriching themselves at the expense of their workers. *Id*.

[7] The Dutch theorist, Jeroen van den Hoven in *Privacy and the Varieties of Informational Wrongdoing*, 27 COMPUTERS & SOC'Y, no. 3, Sept. 1997, at 33, *reprinted in* READINGS IN CYBER ETHICS 493 (Richard A. Spinello & Herman T. Tavani eds. 2004), coined a similar theory of "information inequality" based on the lack of transparency of data brokers' automation and collection efforts, which was further discussed by Nate Cullerton in the context of data collection and credit scoring. *See* Nate Cullerton, *Behavioral Credit Scoring*, 101 GEO. L. J. 907, 819-20 (2013). We expand on this concept, bringing into focus not only the potential discriminatory uses of opaque data collection, but also the lack of privacy regulation placed on data brokers combined with the government's practice of declining to release public data based on extreme privacy concerns, creating data manipulation and destructive rent-seeking behavior.

[8] *See What researchers mean by …primary data and secondary data*, INSTITUTE FOR WORK & HEALTH available at http://www.iwh.on.ca/wrmb/primary-data-and-secondary-data (last visited April 27, 2017) (primary data are collected directly from individual participants in a research project; secondary data is collected from governmental and other public data sources). *See also*, J.H. Reichman, Paul F. Uhlir, *A Contractually Reconstructed Research Commons For Scientific Data In A Highly Protectionist Intellectual Property Environment*, 66-SPG LAW & CONTEMP. PROBS. 315, 354 (2003) (discussing sources of research and data brokers and how the government aids research through public funding of public research).

[9] *See* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701, 1729-30 (2010).

[10] *See generally*, Jillian Raines, *The Digital Accountability and Transparency Act of 2011 (Data): Using Open Data Principles to Revamp Spending Transparency Legislation*, 57 N.Y.L. SCH. L. REV. 313, 344 (2012/2013) (discussing the federal legislation designed to inform the public about tracking federal spending).

[11] *See generally*, Kelsey L. Zottnick, *Secondary Data: A Primary Concern*, 18 VAND. J. ENT. & TECH. L. 193, 196-97 (2015). *See also* Reichman & Uhlir *supra* note 6.

[12] *See* Faith Ramirez et al, *Data Brokers: A Call for Transparency and Accountability*, FTC REPORT, 2014 WL 2217952 (May 2014) *available at* https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf [hereinafter the "2014 Data Broker Report"].

[13] *Id.*

whose data they sold, from learning how the data was gathered based on their trade secret protections.[14] Arguably, this constitutes a form of negative rent-seeking.

In this Article, we examine data brokers' rent-seeking behavior and the legal obstacles posed by potential constraints on this behavior. Next, we discuss the public's right to be informed about the internal workings of the government and about the data it maintains along with the current trend in a "privacy-first" philosophy, constraining governmental officials in their discretionary release of data that should otherwise be considered public. Section IV of our analysis explores the advent of "big data" and complications for researchers in accessing opaque commercial data compared to government open records and privacy laws. In Section V, we address the lack of regulation in the area of privacy obligations and data brokers. Expanding on this topic, we discuss the opaque nature of how data brokers gather and assimilate data in a manner that restricts corroboration and can lead to intentional or unintentional data manipulation, potentially altering the accuracy of research results. Finally, we assert that fewer privacy protections and more data sharing incentives should exist between data brokers and the government or general public to avoid the furtherance of "data inequality." Arguably, it is easier to purchase detailed data about a population than it is to request it from the government, and the purchased data cannot be further examined or corroborated because of the data brokers' intellectual property protections. Naturally, those well-funded researchers or those in collaboration with data brokers will have more opportunities to publish research than less well-funded researchers or the general public. While we acknowledge that this has always been the case, a further imbalance in accessing data that cannot be corroborated will lead to a select few in control of a significant majority of research publications, creating negative rent-seeking and data inequality. We conclude data brokers must share data with researchers and the government to further the public welfare without trade secret limitations and the government must be more flexible in disclosing data, particularly where individuals likely have already placed this data into data brokers' hands through their own use of the Internet. Finally, we recommend that the data brokers encourage transparency for their underlying research through a self-regulatory incentive similar to a "Fairtrade" designation for consumer products. The trademark of "Transparent Data" could be assigned to those entities who willingly allow the underlying methodology of their data to be corroborated and challenged by researchers. In those instances where a data broker wishes to retain the data's secrecy, the data and any research based on such data could include a disclaimer indicating trade secret protection has been asserted, e.g. "Data utilized or provided herein is protected by the providers' intellectual property rights and is not subject to corroboration."

II. Data Brokers Contribute to Negative Rent-Seeking Behavior

Data brokers' use of the governmentally-created Internet combined with their profiteering from freely provided information can be analogized to a form of negative rent-seeking. Originally, the United States government created the Internet as a military tool to collect, store and decentralize data.[15] Today, a nonprofit entity manages the technical aspects of the Internet by assigning

[14] *Id.*

[15] *See* Kim Ann Zimmermann, *Internet History Timeline: ARPANET to the World Wide Web*, (June 4, 2012 12:22 p.m. ET) available at www.livescience.com/20727-internet-history.html. In 1962, the Department of Defense created the precursor to the Internet for military use. *Id.* Additionally, the National Science Foundation maintained control of the Internet hardware. *Id.* Over time, these governmental agencies outsourced their obligations and the

connectivity and root management through a public-private contract.[16] Over time, the government allowed commercial entities access to the Internet and abstained from its control.[17] Commercial entities began utilizing the Internet, and by 2017, the big data industry is expected to reach $47 billion in profits.[18] Data brokers have taken over the data field initially created by the government and supplied by its users.[19] These data brokers sell data that users freely input through governmentally-funded technology.[20] Likewise, with today's technology, there is no meaningful consent to the use of our private data. Users do not understand the extent to which their data can be aggregated and further used to extract their financial expenditures.[21] Users of social media platforms, apps, shopping sites, and who register their email with stores they frequent, are not advised that data are collected by each entity and then sold or given to third parties who then aggregate the data about that person and market their goods, services, or political ideas back to them based on the data they freely provided.[22] Accordingly, data brokers' profits are based in significant part on technology and information created by tax dollars and the general public without a concomitant privacy or disclosure obligations.

By examining the source of a data brokers' underlying business, rent-seeking is apparent. Rent-seeking is a theory of economic behavior that entails asking the government for certain privileges or deriving significant profits and advantages without adding any value to the economy.[23] More simply, it consists of transferring wealth rather than creating wealth.[24] This behavior is criticized as contributing to economic inefficiency and economic inequality as the wealthy receive the benefits of the anticompetitive rent-seeking behavior while the rest of the market suffers the losses.[25] Rent-seeking as an economic theory has been discussed for decades, and involves special interest and lobbying efforts in the form of taxes, subsidies, and preferential regulation in favor of big businesses that have a "symbiotic" relationship with the government.[26]

---

federal government began allowing private commercial entities access to the Internet resulting in the creation of Google and the world wide web. *See* Victoria D. Baranetsky, *Social Media and the Internet: A Story of Privatization*, 35 PACE LAW REV. 304 (2014).

[16] *See* Rolf H. Weber, Shawn Gunnarson, *A Constitutional Solution for Internet Governance*, 14 COLUM. SCI. & TECH. L. REV. 1, 7-8 (2013).

[17] *Id.*

[18] *See* Linda K. Breggin, Judith Amsalem, *Big Data and the Environment: A Survey of Initiatives & Observations Moving Forward*, 44 ENVTL. L. REP. NEWS ANALYSIS 10984, 10986 (2014).

[19] *Id.*

[20] The Department of Commerce predicted that private sector profit from government data ranges from $24-221 billion per year. *See* Frederick Zuiderveen Borgesius, Jonathan Gray, Mireille Van Eechoud, *Open Data Privacy, and Fair Data Principles: Toward a Balancing Framework*, 30 BERK. TECH. L. J. 2073, 2082 (2015).

[21] *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1886 (2013) (describing consumers' lack of knowledge regarding the amount of their data that is used by private data brokers).

[22] *Id.*

[23] *See* Mark Seidenfeld, Murat C. Mungan, *Duress as Rent-Seeking*, 99 MINN. L. REV. 1423, 1426 & n. 18 (2015). *See also* Tankersley, *supra* note 2.

[24] *See* Seidenfeld et al. *supra* note 21 at 1437.

[25] *See* Joseph P. Tomain, *Gridlock, Lobbying, and Democracy*, 7 WAKE FOREST J. OF LAW AND POLICY 87, 101 & 110 (2017).

[26] *See* Todd J. Zywicki, *Rent-Seeking, Crony Capitalism, and the Crony Constitution*, 23 SUP. CT. ECON. REV. 77, 78-79 (2016) (citing Mancu Olson's 1982 work *The Rise and Decline of Nations* which addressed how interest groups capitalize on their power and influence over legislators to obtain special favors). A common illustrative of

We contend, similar to legal scholars who have likened rent-seeking to contractual duress,[27] that data brokers' reuse of a person's aggregated data without their clear affirmative consent is a form of inappropriate rent-seeking, leading ultimately to further economic and data inequality. Where the contracting party forces, via economic or other duress, another party to consent to contract where he might not otherwise have done so, the threat-maker can be said to be engaged in detrimental rent-seeking behavior through the transfer of wealth without coordinate value provided by the threat-maker.[28] At its core, rent-seeking can be seen as the commercialization of existing resources without the input of additional value, as opposed to profit-seeking which embodies mutually beneficial transactions.[29] Where efforts are rewarded in the form of wealth redistribution from the rent-seeking behavior rather than by earning wealth through productive activity, the rent-seeking behavior will increase, taking away productive jobs and creativity from our society.[30] By way of example, the technology industry has utilized various forms of rent-seeking behavior to gain regulatory advantages for their fields as well as to drive their competitors out of business. Recently, Microsoft lobbied the Federal Trade Commission ("FTC") to investigate Google, its competitor, for antitrust violations for over two (2) years, costing Google approximately $25 million in counter-lobbying efforts.[31] Ultimately, Google succeeded in averting an antitrust suit, which some have hypothesized was due to its frequent access to the Obama White House and key decision-makers.[32] In this regard, rent-seeking

---

illegal rent-seeking behavior's societal costs is the common criminal. *Id*. at 80-81. The criminal forgoes other productive activity, including gainful employment, and diverts third parties' resources, causing them to purchase theft insurance, alarms, etc., rather than engaging in otherwise productive endeavors and purchases. *Id*. Another common example of negative rent-seeking is the rate at which capital gains taxes are calculated. *See generally*, Joseph E. Stiglitz, *A Tax System Stacked Against the 99 Percent*, (April 14, 2013 9:36 p.m. ET) available at https://opinionator.blogs.nytimes.com/2013/04/14/a-tax-system-stacked-against-the-99-percent/?_r=0. Many of the country's wealthiest individuals are paying taxes only on their carried interest (their passive investments) rather than on actively earned income (because they are not engaged in active employment). *Id*. This is the rent-seeking aspect of their profits, and man legislators have called for reform, requiring the carried interest income be taxed at the individual's ordinary income rate to avoid the consequences of negative rent-seeking behavior. *See generally*, Joseph E. Stiglitz, *A Tax System Stacked Against the 99 Percent*, (April 14, 2013 9:36 p.m. ET) available at https://opinionator.blogs.nytimes.com/2013/04/14/a-tax-system-stacked-against-the-99-percent/?_r=0 and Michael Cragg & Rand Ghayad, *Inequalities in Tax Policy* (May 4, 2015 7:30 p.m. ET) available at http://www.huffingtonpost.com/rand-ghayad/inequities-in-tax-policy_b_7209108.html.

[27] *See* Seidenfeld et al. *supra* note 21 at 1437. *See also* Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 549 (2014). However, there is a countervailing argument that data mining adds value by collecting, sifting, consolidating and distributing data in new ways that would otherwise be prohibitively time consuming. *See* Ruth L. Okediji, *Government as Owner of Intellectual Property? Considerations for Public Welfare in the Era of Big Data*, 18 VAND. J. ENT. & TECH. 331, 335 (2016).

[28] *Id*.

[29] Big data have special value and "resides downstream from commercial exchanges that take place between data producers and their consumers." *See* Mattioli, *supra* note 25 at 549. However, there is a countervailing argument that data mining adds value by collecting, sifting, consolidating and distributing data in new ways that would otherwise be prohibitively time consuming. *See* Okediji, *supra* note 25 at 335.

[30] *See generally*, *id*.

[31] *See* Zywicki *supra* note 24 at 84.

[32] *See* Johnny Kampis, *Visitor Logs Show Google's Unrivaled White House Access* (May 16, 2016) available at http://watchdog.org/265252/visitor-logs-google-white-house.

behavior and one's successful lobbying, rather than one's successful market venture, dictates an entity's economic survival.[33]

Data brokers and others will likely contend that they are not rent-seekers as they provide added value by aggregating discrete data sets through specific algorithms which they have independently created allowing third parties to develop a fuller picture about their consumers and providing consumers with information of specific interest to them.[34] These algorithms are the data brokers' protected trade secrets, to which the general public does not have access despite their significant contribution to the data brokers' profit.[35] The aggregated data is sold to third parties for marketing purposes.[36] In this regard, data brokers simply access free data from individuals who are generally unaware that their data are being repurposed and, then resell it to third parties who wish to market to the same individuals and obtain further purchases from them.[37] Although entirely legal, the combination of these three (3) factors: (1) individuals freely providing data without realizing its resale value, (2) legally protected aggregation of individuals' data, and (3) use of this data to convince the users to purchase products they might not have otherwise bought, arguably is the transference of wealth rather than the creation of wealth.[38] We contend this is a form of detrimental rent-seeking in need of reform.

A. Constitutional Obstacles in Controlling Rent-Seeking Behavior

Despite the negative consequences to our economy from rent-seeking behavior, an obstacle to its constraint is the United States Constitution. Data brokers' aggregation and resale of data is commercial speech protected by the First Amendment.[39] In *Citizens United v. Federal Election*

---

[33] *See* Zywicki, *supra* note 24 at 102-03. Although lobbying or other forms of political rent-seeking are legal, as opposed to illegal rent-seeking like a thief, the economic effects are similar. *Id*. at 80-82. The opportunity costs associated with seeking governmental privileges include the travel costs for the corporate officers to meet with politicians, their campaign contributions, and money spent influencing officials, rather than the corporate officers' utilizing their time to manage their business in a more efficient manner. *Id*.

[34] *See* Okediji, *supra* note 25. *See How Much Is Your Personal Data Worth*? available at https://www.webpagefx.com/blog/general/what-are-data-brokers-and-what-is-your-data-worth-infographic/ (detailing how data brokers often obtain data from you and how much some users are paid for their data, noting the majority of data comes from public data, or consumers' voluntary input of data into a variety of sources like loyalty programs).

[35] *Id. See also* Adam M. Samaha, *Government Secrets, Constitutional Law, and Platforms for Judicial Intervention*, 53 UCLA L. REV. 909, 922 (2006) "Openness exposes not just waste, fraud, and abuse, but also ... candid advice, intimately private data, and trade secrets" and "[a] rule of full disclosure might also prompt officials to sanitize the public record as it is created").

[36] *See* Okediji *supra* note 25.

[37] Scholars may contend that data and marketing materials directed to our special interests are likewise added value and not rent seeking. *Id*.

[38] More plainly, data brokers use technology which was created at tax payer expense, the Internet, to repurpose information which users freely and unwittingly provide to one source. Users are unaware that the information they have provided will be combined with other information they provide to different sources online and then resold. Users are not provided meaningful disclosure of this aggregation and reuse nor are they generally compensated. Likewise, data brokers pay nothing to the government (outside their donations or taxes) for the use of the Internet, and while some may argue that the Internet is a free public resource, the sheer amount of profit derived from the Internet and freely provided information warrants some additional consideration, whether it be through disclosure obligations or opt-out mechanisms described throughout this article.

[39] *See* Richard L. Hasen, *Lobbying, Rent-Seeking, and the Constitution*, 64 STAN. L. REV. 191, 198-99 (2012).

Com'n, the United States Supreme Court held that corporations are people in the eyes of the First Amendment and have the right to support a political viewpoint and candidate through financial means, rendering legislation on lobbying efforts difficult, although not impossible.[40] The Supreme Court noted that the government could "regulate corporate political speech through disclaimer and disclosure requirements, but it may not suppress that speech altogether."[41] With respect to political rather than commercial speech, the *Citizens United* court found restrictions on political speech "are 'subject to strict scrutiny,' which requires the Government to prove that the restriction 'furthers a compelling interest and is narrowly tailored to achieve that interest.'"[42] Regarding the underlying disclaimer and disclosure requirements of the challenged law, the Court determined these provisions are governed under "exacting scrutiny" which requires a "substantial relation" between the disclosure requirement and a "sufficiently important" governmental interest.[43] The Supreme Court found that disclosure requirements are a "less restrictive alternative to more comprehensive regulations of speech."[44] In *Citizens United*, the Court found no evidence that the disclosure requirements as applied would expose the individuals identified by the disclosure to harassment or abuse, and were constitutional.[45] The Court rationalized that the disclosure obligations allow the electorate to be fully informed and give proper weight to the speakers and their messages.[46] Because the essence of the First Amendment is to ensure free and open discourse, the disclosure obligations further an important interest, and were found constitutional.[47]

Along these lines, legal scholars have argued that national economic welfare and the idea of income or social inequality are compelling interests warranting narrowly tailored regulations of commercial or political speech.[48] Ultimately reducing rent-seeking behavior improves economic productivity and can lead to a decrease in the deficit, while permitting rent-seeking behavior leads to slow, long-term economic growth.[49] Reducing the nation's growing income inequality resulting from the diversion of wealth by unproductive means and anticompetitive behavior are arguably compelling governmental interests that would withstand even the most exacting judicial scrutiny espoused by *Citizens United*.[50] Accordingly, specific legislation designed to reduce a data brokers' rent-seeking behavior should be constitutional where designed as a disclosure

[40] *Id*. *See also* Citizens United v. Federal Election Com'n, 558 U.S. 310, 342 (2010).
[41] *See Citizens United,* 558 U.S. at 318.
[42] *Id*. at 340 (internal citations and quotations omitted).
[43] *Id*. at 366 (finding that although disclaimer and disclosure requirements can "burden the ability to speak"… "they 'do not prevent anyone from speaking[.]'") (internal citations omitted)
[44] *Id*. at 369.
[45] *Id*. at 370. There may be times when a disclosure obligation as applied is unconstitutional, e.g. when there is a "reasonable probability that the group's members would face threats, harassment, or reprisals if their names were disclosed." *Id*.
[46] *Id*. at 371.
[47] *Id*. at 371. However, the Court did find limitations on corporate lobbying and political contributions unconstitutional. *Id*. at 340.
[48] *See* Hasen *supra* note 37 at 198-99. The level of judicial scrutiny depends on the type of behavior being regulated. For example, campaign contributions receive a lower form of First Amendment protection because they are a form of commercial speech as opposed to bans on political speech, like solicitation prohibitions and revolving door statutes that are subject to strict scrutiny. *Id.* at 239.
[49] *Id*. at 232 (discussing economists' studies on economic market growth in eras with high and low rent-seeking activity and noting a significant correlation with negative results in high rent-seeking eras).
[50] *See* Tomain, *supra* note 23 at 113. *See also Citizens United*, 558 U.S. at 318.

obligation to advise users of the full extent of their use, aggregation and resale of their information and whether research results are based on the ability to corroborate the underlying research data. A more restrictive obligation requiring data brokers to actually provide and share their underlying data with the government or researchers without charge so that the data can be used, corroborated and challenged would likely be unconstitutional unless strictly circumscribed to very limited situations, e.g. information related to significant public welfare or national security issues.[51]

The inability to curb a data brokers' anticompetitive behavior based on First Amendment grounds could be a significant hurdle,[52] and one that will force legislators to address regulations in terms of national economic welfare and inequality, including the compelling reasons behind such legislation. A prime example of a defeated attempt to constrain commercial data brokers' speech is evidenced by the Supreme Court's ruling in *Sorrell v. IMS Health, Inc*.[53] IMS is a data broker in the health care field and analyzes trends for certain companies who, in turn, market to their customers.[54] One service IMS provides is the service of "detailing" in which IMS collects prescription and purchasing data from individual pharmacies, identifying physician trends in prescribing pharmaceuticals and then profiling the physicians in an effort to assist pharmaceutical companies in marketing their drug to those doctors.[55] Two states, Vermont and New England, sought to lower the cost of prescription drugs by restricting data detailing because the costs of such detailing were passed on to consumers.[56] The laws prohibited pharmaceutical and insurance companies from selling prescription data to the data brokers.[57] IMS and others challenged the laws on corporate free speech grounds, and the Supreme Court ultimately ruled the laws unconstitutional.[58] The Court found the statutes were content-based restrictions on the marketer and the data's use rather than a ban on other forms of speech on the same topic.[59] The Court went on to note: "the creation and dissemination of data are speech within the meaning of the First Amendment."[60]

Since *Sorrell*, courts have determined that despite *Sorrell's* proclamation that disclosure of consumers' identities is protected speech, legislative regulation of speech may still be

[51] An outright requirement of free provision of information could be construed as an unconstitutional taking, U.S. Const. Amend. V, or conversion of a data brokers' property. *See* Mark A. Lemley, *Private Property*, 52 STAN. L. REV. 1545, 1549 (2000) (allowing individuals to have a property right in the data they contribute online would have a negative impact on commerce). *See generally*, Mark Bartholomew, *Intellectual Property's Lessons For Information Privacy*, 92 NEB. L. REV. 746, 756 (2014) (discussing the trend that data collectors have nearly a undisturbed right to free speech and proposing a balancing analysis for courts). *See e.g*., Trans Union LLC v. Credit Research, Inc., 2001 WL 648953 (N.D.Ill. June 4, 2001) (holding possible conversion claims where a licensee exceeds the use agreement for on-line data) and Tahoe-Sierra Preservation Council, Inc. v. Tahoe Regional Planning Agency, 535 U.S. 302 (2002).
[52] *See* Tomain, *supra* note 23 at 113.
[53] 131 S. Ct. 2653 (2011).
[54] *Id*. at 2660.
[55] *Id*.
[56] *Id*. at 2668.
[57] *Id*. at 2663.
[58] *Id*. at 2661-62.
[59] *Id*. at 2663-64 & 2667.
[60] *Id*. at 2667.

appropriate.[61] Speech that is commercial in nature, i.e. marketing goods and services, is subject to less than strict scrutiny.[62] As a result, legal scholars have advocated for more regulation in the area of data privacy to include providing consumers with the right to affirmatively consent to the data that are gathered about them and to have the ability to correct that data combined with better privacy protections on the types of data a data broker may sell and the security protections they utilize in maintaining this data.[63] Whether such regulation would violate *Sorrell* and the First Amendment is the subject of debate.[64] As discussed more thoroughly below in Section V, privacy laws generally do not protect individuals from data brokers' resale of their data.[65] Rather, data brokers and entities rely on both the consumers' acceptance of their terms of use and their broad privacy policies to reuse and repurpose the consumer's data.[66]

---

[61] *See* Boelter v. Hearst Communications, Inc., 192 F. Supp. 3d 427, 444 (S.D.N.Y. 2016). Lower courts have struggled with the implications of *Sorrell* in the context of compelled disclosures surrounding commercial speech. *See* Note, The Harvard Law Review Association, *Repackaging Zauderer*, 130 HARV. L. REV., 972, 979-983 (2017). The Sixth Circuit requires the lesser rational basis scrutiny for governmental restrictions on commercial speech that is "likely" to mislead; whereas, the Fifth and Eighth Circuits allow a rational basis test where the commercial speech is "potentially misleading." *Id.* at 980. *See also* Zauderer v. Office of Disciplinary Counsel, 471 U.S. 626, 651 (1985) (holding mandatory disclosure obligation of "purely factual and uncontroversial information" aimed at preventing consumer deception does not violate the First Amendment.).

[62] *See* Boelter v. Hearst Commc'ns, Inc., 192 F. Supp. 3d at 447. One area where data brokers profit from rent-seeking behavior is in the aggregation of existing medical data that individuals willingly provide to healthcare professionals, to investigative studies, or to other online services. *See* Adam Turner, *How Data Brokers Make Money Off Your Medical Records* (Feb. 1, 2016) *available at* https://www.scientificamerican.com/article/how-data-brokers-make-money-off-your-medical-records/. With current technology, there is a greater risk that a person's anonymized data can be reverse engineered to reveal his identity. *Id.* The most significant data broker in this market is IMS Health, Inc. ("IMS") which recorded $2.6 billion in revenue in 2014. *Id.* Pfizer, the pharmaceutical giant, pays $12 million to purchase this type of data from data brokers, including IMS. *Id.*

[63] The United States Department of Health, Education and Welfare developed privacy guidance documents. *See* Borgesius et al. *supra* note 18 at 2109. One of the guidelines is to restrict repurposing data collected for other reasons without consent. *Id.*

[64] *See* Neil M. Richards, *Why Data Privacy Law is (Mostly) Constitutional*, 56 WILLIAM AND MARY L. REV. 1501, 1521 (2015).

[65] Data is collected through a person's browsing history, through a person's purchases, and by tracking their cookies. *See Big Data A Tool for Inclusion or Exclusion*, FTC REPORT, 2016 WL 23163 at *7-8 (Jan. 2016) (detailing how data is gathered, analyzed and used). Although the Health Insurance Portability and Accountability Act ("HIPAA") protects individual's private medical information, it is inapplicable to data brokers' marketing and research activities. *See* Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.) and *see* 2014 Data Broker Report *supra* note 10. Nonetheless, such information can be used or disclosed when certain anonymization techniques eliminate the ability to identify an individual, such as generalizing birth dates, zip codes, etc. *Id.* Data brokers are generally exempt from HIPAA obligations because data brokers receive or purchase aggregated and de-identified data from a covered entity, meaning the individual's identity is not provided to the data broker. *Id. See also Covered Entities*, HIPAA PRIVACY RULES, *available at* http://www.hhs.gov/ocr/privacy/hipaa/understanding/coveredentities/. *See also* "*To Whom Does the Privacy Rule Apply and Whom Will It Affect?*" U.S. DEPART. OF HEALTH AND HUMAN SERVICES (Sept. 20, 2015 8:11 p.m.) *available at* http://perma.cc/N3FJ-57P7. The data brokers can aggregate the data along with statistical, de-identified data gathered from pharmacies and insurance companies, creating a valuable data commodity that third parties purchase to educate them on future investments, and marketing schemes. *Id.* Although personally identifiable health data are technically confidential and only statistical data about individuals are gathered and aggregated, the reality is that a third party can discover a person's identity from this data. *Id.*

[66] *See* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1880, 1886 (2013) (describing consumers' lack of knowledge regarding the amount of their data that is used by private data brokers).

With these two Supreme Court cases as a backdrop, any governmental restrictions on data brokers' powers to collect and sell data must withstand significant judicial scrutiny. The manner in which regulations could withstand this scrutiny lies within the realm of disclosure requirements rather than restriction requirements.[67] It is within this area that we contend the lack of regulation will lead to data inequality, and arguably violate the public's right to access data violating the public's First Amendment rights.[68]

III. Modern Complications for a Researcher's Access to Data

Data brokers control the nature and access to data which will form the basis of future research. Many data brokers and other technology companies collaborate with researchers to conduct a variety of research, but the underlying data and processes surrounding their research are never made public.[69] In this regard, the underlying data cannot be checked or challenged.[70] If access to similar data is unavailable without purchase from the data broker, then necessarily only the wealthiest researchers or only those with special relationships with data brokers will have their voices heard, detrimentally impacting our national public welfare. Accordingly, data brokers should be required to share certain big data necessary for public research with the government, a form of disclosure requirement subject to the Supreme Court analysis noted above.

Big data has altered human subject research, including expanding the definition of who is a researcher.[71] Innovative application developers and others collect significant amounts of data without oversight.[72] The flexible practices and the public's inability to test the researchers' analysis of private data can lead to a "credibility crisis in computational science, not only among scientists, but as a basis for policy decisions and in the public mind."[73] Scholars agree that the lack of transparency regarding what data are used, how data are collected, and how data are analyzed are significant issues with new research.[74] In this regard, access to public data is paramount to the future of credible and unbiased research. Notably, the public obtains data from two (2) primary sources: private data brokers, social media, search engines, app providers, and

---

[67] *See* Richards, *supra* note 62 at 1521.

[68] *See* Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 583 (1980) (holding where government creates an arbitrary obstacle to important data, it is violating the person's, who is trying to access the data, First Amendment rights).

[69] *See e.g. Facebook Partners with 17 Universities to Streamline Research* REUTERS (Dec. 21, 2016) *available at* http://fortune.com/2016/12/21/facebook-universities-research/. *See also* Jaikumar Vijayan, *Google Invites University Researchers to Collaborate on IoT Projects* (Feb. 12, 2016) *available at* http://www.eweek.com/networking/google-invites-university-researchers-to-collaborate-on-iot-projects.

[70] *See e.g.*, Reichman & Uhlir, *supra* note 6 at 319-320 & 354 (analyzing the nature of scientific data and the advent of data brokers and their desire to protect their research outcomes through intellectual property mechanisms and recommending additional university and governmental public research).

[71] *See* Omer Tene, Jules Polonetsky, *Judged by the Tin Man: Individual Rights in the Age of Big Data*, 11 J. TELECOMM. & HIGH TECH. L. 351, 353 (2013).

[72] *Id.* at 353.

[73] *Id*. at 354 (internal citation omitted).

[74] *Id.* at 355.

the United States government.[75] The United States government provides data to the public in three (3) general forms: (1) through the general release of mass data, (2) upon request for records maintained by the government through requests under the Freedom of Information Act ("FOIA"), and (3) through governmentally funded research.[76]

A. What is Data?

Data are defined as the "representation of facts or ideas in a formalized manner capable of being communicated or manipulated by some process."[77] Datafication is the "act of rendering these representations into a format that can be communicated or manipulated by some process."[78] Researchers often utilize a combination of sources including public data and commercial data.[79] Further, agencies release data through a variety of means, including: "derived index data, aggregated tables or sanitized microdata in public use data files, raw data controlled via a secure data enclave, or, to a lesser extent, data made available online through query systems."[80] The term "big data" includes the "novel ways in which organizations including the government and businesses combine diverse digital datasets and then use statistics and other datamining techniques to extract from them both hidden and surprising correlation."[81] Big data begin in the form of small segments of data collected, consolidated and analyzed.[82] Entities like advertising networks, social media, banks, and retailers analyze the data and build consumer profiles, storing billions of data elements on consumers and then predict how the consumer will behave based on the profile.[83] Data mining is the complex process of taking data collected from a variety of sources, both public and private, removing unreliable or redundant data, and constructing statistical models using the remaining data such that anyone in possession of the mined data can predict future behaviors.[84] It is this aggregation that data brokers wish to protect and would assert adds value to the economy.[85] However, the added value is at the extreme expense of unwitting users and the research community, and at inordinate profit to the data broker who

---

[75] *See generally*, Jennifer Bresnahan, *Personalization, Privacy, and the First Amendment: A Look at the Law and Policy Behind Electronic Databases*, 5 VA. J. L. & TECH. 8 (2000) (discussing the strong constitutional protections for data brokers, their databases, and their data mining practices).

[76] *See* Micah Altman, Alexandra Wood, David O'Brien, Salil Vadhan, Urs Gasser, *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 BERKELEY TECH. L. J. 1967, 1991 (2015) and Reichman & Uhlir *supra* note 6.

[77] *See* Meg Leta Ambrose, *Lessons From the Avalanche of Numbers Big Data in Historical Perspective*, 11 I/S: J.L. & POLICY FOR INFO. SOC'Y 201, 210 (2015) (internal citation and quotation omitted).

[78] *Id.* at 211.

[79] *See* Altman et al *supra* note 74 at 2001. Unlike public data sets, restricted data requires researchers to apply for access to the data, and the governmental release depends on a formal screening process. *Id*. at 1996. The use is limited to the purposes specified through data use agreements. *Id*. at 1996.

[80] *Id.* at 1993.

[81] *See* Ambrose *supra* note 75 at 212. Big data refers to "a new method of empirical inquiry." *See also*, Mattioli *supra* note 25 at 539.

[82] *See* Mattioli *supra* note 25 at 539.

[83] *Id.*

[84] *See* Zottnick, *supra* note 10.

[85] *See* J.H. Reichman, Paul F. Uhlir *supra* note 63 at 354 & 368-69.

would not exist without the governmentally created Internet and without users providing them with free information.[86]

Every second there are approximately 6,000 tweets, 40,000 Google searches, and individuals send over 2 million emails.[87] As of 2014, there were 1 billion websites on the Internet.[88] There are generally three (3) types of relevant research data: (i) aggregated data (summary information released to the public),[89] (ii) de-identified microdata released to researchers for analytical purposes ("data [released] in its most granular, unaggregated form."),[90] and, (iii) identified data (customer identification, at least in some form like an IP address, necessary for targeted marketing purposes).[91] The importance of big data and a data broker's role in today's research cannot be overstated.

B. The Importance of Publicly Available Data

Historically, the general public and academic researchers relied on data gathered and disseminated by "public institutions" (including government agencies, non-profit organizations, universities, research centers, and others, by accessing routinely publicized agency data releases).[92] Public data or the "data commons" inform the public and enhance research.[93] Increased demand for privacy in recent years has led government agencies to be less inclined to share their data or to enact data protection measures that diminish the utility of the data.[94] Simultaneously, there has been an exponential increase in the quantity and quality of data collected by private sources as discussed above.[95] These private sources can purchase the collected data with or without disclosing individual identities.[96]

---

[86] *Id.* at 371.

[87] *See* Stephanie Pappas, *How Big is the Internet, Really?* (March 18, 2014 11:40 a.m. ET) *available at* www.livescience.com/54094-how-big-is-the-internet.html.

[88] *Id.*

[89] *See* Briefing Paper on Open Data and Privacy, THE CENTER FOR OPEN DATA ENTERPRISE *available at* http://reports.opendataenterprise.org/BriefingPaperonOpenDataandPrivacy.pdf (last visited April 3, 2017).

[90] *Id.*

[91] *Id*.

[92] *See* authorities discussed *supra* note 6.

[93] Re-use of public data creates new business, services, and productivity. *See* Farhnem Jahanian, *Policy Infrastructure for Big Data: From Data to Knowledge to Action*, 10 I/S J. L. & POL'Y FOR INFO. SOC'Y 865, 866-868 (2015). For example, financial service providers use statistics for input, and the meteorological field uses weather data to provide specific forecasting for off-shore oil companies. *Id*. *See* Borgesius et al. *supra* note 18 at 2081.

[94] *See generally*, J. Trent Alexander, Michael Davern & Betsey Stevenson, *Inaccurate Age and Sex Data in the Census* PUMS *Files Evidence and Implications* 1-3 (CESifo Working Paper No. 2929, 2010), *available at* http://ssrn.com/abstract=1546969; *See* Steven Levitt, *Can You Trust Census Data?*, FREAKONOMICS (Feb. 2, 2010, 11:09 AM) *available at* http://www.freakonomics.com/2010/02/02/can-you-trust-census-data. *See* Lara Cleveland, Robert McCaa, Steven Ruggles, Matthew Sobek, *When Excessive Perturbation Goes Wrong and Why IPUMS-International Relies Instead on Sampling, Suppression, Swapping, and Other Minimally Harmful Methods to Protect Privacy of Census Microdata*, (2012) *available at* http://link.springer.com/chapter/10.1007%2F978-3-642-33627-0_14.

[95] *See* Joseph A. Tomain, *Online Privacy & the First Amendment: An Opt-In Approach to Data Processing*, 83 U. CIN. L. REV. 1, 3 (2014).

[96] *Id.*

Arguably, the increase in demand for privacy is driven primarily by private sources collecting and selling the data; yet, there are minimal, if any, private data aggregator privacy requirements (legal or otherwise).[97] In most cases, the private sources rely on the user's "consent" (usually in the form of clicking a box when using a website, downloading an application, etc.).[98] The very availability of data from these private sources has led some to demand that the *government* enhance its data protection to prevent those with access to the private sources of data from potentially reverse engineering the individual's identity before accessing the government's records.[99] To us, this seems unfair forcing the government to protect individual privacy while private data brokers bear limited similar burdens.

IV. Open Access to Data within the Government's Control

There has been considerable research on the practices of government agencies both from a technical and policy perspective.[100] But to our knowledge, the effect of increased governmental privacy obligations on researchers and the public has not been examined. Central to a democratic environment is a philosophy of the citizenry's right to know about the internal workings and decisions of its government and the data that it maintains.[101] Although our colonists believed in the idea of a public's right to know, the United States Constitution does not contain such a provision, and in fact, the founding fathers were less than transparent in their management of the government.[102] It was not until 1943, in *Martin v. City of Struthers, Ohio*, that the United States Supreme Court first recognized a constitutional right to *receive* data under the First Amendment.[103] Thereafter by the 1940s, many states enacted legislation governing the retention and maintenance of government records, and the federal government enacted the Administrative Procedures Act, establishing internal operating and records retention procedures for federal agencies.[104] World War II was being waged during this timeframe, and many secrecy and censorship laws were in place for national security reasons.[105] After the end of the war, President Harry S. Truman continued classifying many records as "secret", drawing significant

---

[97] *See* Drivers Privacy Protection Act, 18 U.S.C. § 2721 et seq. (1994 & Supp. 1998); the Video Privacy Protection Act, 18 U.S.C § 2710 (2013); Children's Online Privacy Act, 15 U.S.C. § 6501 (2000); Federal Credit Reporting Act, 15 U.S.C. § 1681 (2006); the Gramm-Leach Bliley Act, 15 U.S.C. § 6801 (2006) and Health Insurance Portability and Accountability Act of 1996, Pub. L. No. 104-191, 110 Stat. 1936 (codified as amended in scattered sections of 18, 26, 29, and 42 U.S.C.). *See also*, Paul M. Schwartz, *Privacy and Democracy in Cyberspace*, 52 VAND. L. REV. 1609, 1611 (1999) (noting lack of standards for cyberspace privacy, legal or otherwise).

[98] *See generally*, Tomain *supra* note 93.

[99] *See* Nancy S. Kim, D. A. Jeremy Telman, *Internet Giants as Quasi-Governmental Actors and the Limits of Contractual Consent*, 80 MO. L. REV. 723, 728-29 (2015). *See also* Anne Klinefelter, *When to Research is to Reveal: The Growing Threat to Attorney and Client Confidentiality From Online Tracking*, 16 VA. J.L. & TECH. 1 (2011).

[100] *See* Andrew Chin, Anne Klinefelter, *Differential Privacy as a Response to the Reidentification Threat: The Facebook Advertiser Case Study*, 90 NC L. REV. 1417, 1427 (2012).

[101] A concept that dates back to the Athenians in 330 B.C. *See* David Cuillier, *The People's Right to Know: Comparing Harold L. Cross' Pre-FOIA World to Post-FOIA Today*, 21 COMMUN. L. & POL'Y 433, 438 (2016).

[102] *Id.* at 439 & n 26 (explaining how the early stages of the U.S. Government acted in secrecy).

[103] *See* 319 U.S. 141, 143 (1943) (holding freedom of speech protections encompass the "right to distribute literature.").

[104] *See* Cuillier *supra* note 99 at 441.

[105] *Id*.

and widespread journalistic criticism.[106] These concerns led to the American Society of News Editors' report, regarding "customs, laws and court decisions affecting our free access to public information whether it is recorded on police blotters or in the files of the national government."[107] FOIA was enacted as a result of this report and is one form of governmental release of records.[108]

Outside the context of FOIA, a second method of government data release is common public sector data publication, including government performance data.[109] Government performance data are defined as data that "can be freely used, modified, and shared by anyone for any purpose."[110] Governments struggle with the benefits of releasing public data and the potential privacy implications.[111] However, several federal agencies routinely release public sector data.[112] For example, the Census Bureau releases statistical data about individuals in an aggregated form gathered from interviews and questionnaires, creating official statistics from tabular or relational data.[113] Voluntary release of data provides for transparent research, as opposed to research protected by intellectual property laws, and enables other researchers to test the original researcher's analysis and opinion for a more thorough examination of the topic.[114] Despite the importance of publicly available data, individual privacy in the data is likewise significant.

A. Open Access to Governmental Data under FOIA vs. Personal Privacy

While access to information is important for researchers, individual privacy interests in protecting sensitive data is likewise paramount. One aspect of the privacy analysis that has received little attention is how privacy demands affect data release associated with FOIA requests.[115] The relationship between the researcher and the government agency is very different

[106] *Id*. at 440.
[107] *Id*. at 441 (quoting Harold L. Cross, *The People's Right to Know: Legal Access to Public Records and Proceedings XIII* (1953).
[108] *See id*. at 442-43 (providing a detailed history of the basis for FOIA which took many of its provisions from excerpts of state laws, common laws, case law, attorney general opinions and agency regulations as half of the states had public disclosure laws).
[109] *See* "Census Data Mapper", UNITED STATES CENSUS BUREAU *retrieved at* census.gov (last visited April 27, 2017).
[110] *See*, Borgesius et al., *supra* note 18 at 2076.
[111] *Id*.
[112] *See* "Census Data Mapper" *supra* note 107.
[113] *See* Altman et al. *supra* note 74 at 1991. Certain government agencies gather and release statistical data to assist government policy and economic decisions, research and transparency. *Id*. at 1991.
[114] *See* Rebecca Lipman, *Online Privacy: The Invisible Market for Our Data*, 120 PENN. ST. L. REV. 777, 790-92 (2016).
[115] Although FOIA and open records laws often speak in terms of FOIA disclosures, research in this field tends to define a disclosure as an unintentional release of sensitive data rather than the voluntary release of data. *See generally* Felix T. Wu, *Defining Privacy and Utility in Data Sets*, 84 UNIV. OF COLO. L. REV. 1117, 1118-1120 (2013). Thus, in this Article the authors utilize the term "release" when data is voluntarily and properly released as

in the context of the voluntary release of data as opposed to FOIA releases. Government agencies voluntarily release data under legal mandates (as in the case of the Census Bureau) or as an integral part of their function.[116] Whereas FOIA requests often place the researcher, and the agency in an adversarial position, i.e. the agency is reluctant to release the data and the FOIA request forces the data's release.[117] FOIA's main goal is to ensure the public is informed about the government so that it can be held accountable for its actions.[118]

Despite the premise of transparency first, there are nine (9) exemptions allowing the government to refuse to release records in the government's possession.[119] Initially, FOIA exemptions were not designed as "mandatory bars to disclosure."[120] Rather, the exemptions provide the agency with *discretion* to withhold data in its possession.[121] The exemptions are written in terms of the government "may" withhold the release of records, instead of "prohibiting" their release.[122] Where data does not fall within one of these exemptions, discretionary government release of data may be permissible and appropriate.[123] However, certain exemptions may be inappropriate for discretionary government release, including Exemption 6 (personnel and medical records) and 7(C) (records containing personal privacy concerns).[124] The lack of clarity regarding when a discretionary release is appropriate, particularly under Exemptions 6 and 7(C) for privacy reasons, causes agencies to err on the side of privacy protection and not release the data.[125] Evidence that agencies decline to release data under FOIA is revealed through a comparison of FOIA release outcomes under both the Bush and Obama administrations.[126] As can be seen from

---

opposed to "disclosure" relating to the release of data that may contain information that leads to de-identification of an individual. *Id*.

[116] *See* "Census Data Mapper" *supra* note 107.

[117] Freedom of Information Act, 5 U.S.C. § 552 (2006), as amended by OPEN Government Act of 2007, 5 U.S.C. § 552 (Supp. II 2008).

[118] *See* John Doe Agency v. John Doe Corp., 493 U.S. 146, 152, *reh'g den.*, 493 U.S. 146 (1989). FOIA specifically provides that the government shall release records "upon any request for records, which (i) reasonably describes such records and (ii) is made in accordance with published rules. . ." 5 U.S.C. § 552(a)(3)(A) (2012).

[119] 5 U.S.C. § 552(b) (2012).

[120] *See* Chrysler Corp. v. Brown, 441 U.S. 281, 293 (1979), *superseded by statute in* Iowa Film Prod. Serv. v. Iowa Dept. of Econ. Dev., 818 N.W.2d 207 (Iowa 2012).

[121] *Id*. at 294.

[122] *See* Max Galka *"Analyzing FOIA Statistical Trends From 2011 to 2012*", FOIA Blog (Aug. 28, 2013) *available at* http://ains.com/foiablog/2013/8/28/analyzing-foia-statistical-trends-from-fy2011-to-fy2012.html.

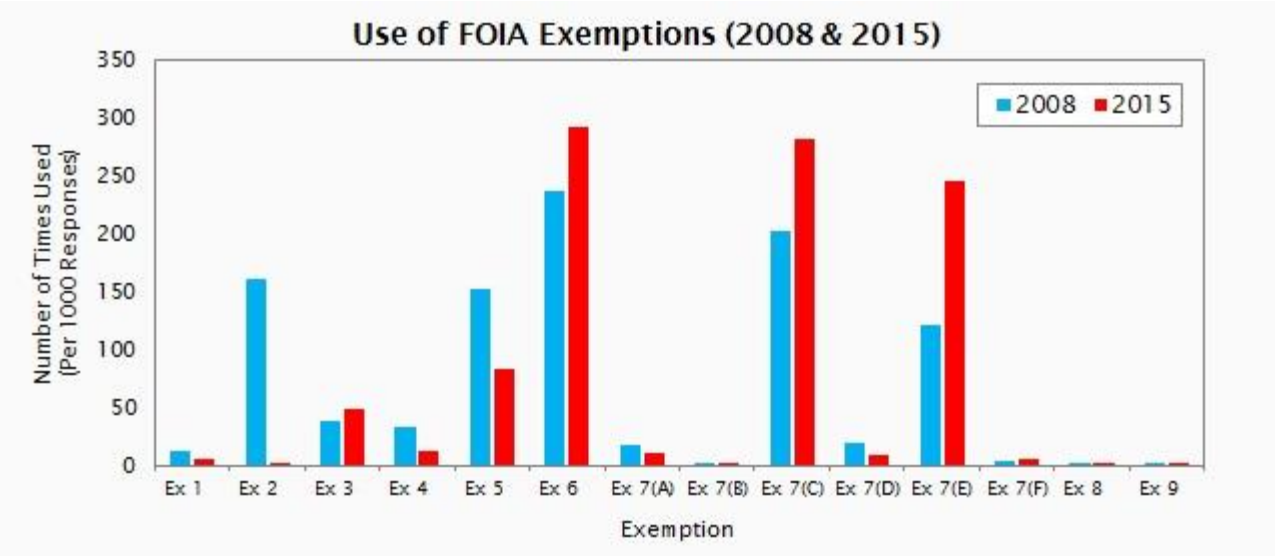
[123] *Id*.

[124] *Id*. *See also* Department of Justice *Guide to the Freedom of Information Act Exemption 6 available at* https://www.justice.gov/sites/default/files/oip/legacy/2014/07/23/exemption6_0.pdf (last visited April 27, 2017).

[125] *See e.g*. Department of Justice, *Data*, *available at* https://www.foia.gov/data.html (last visited April 27, 2017) (by selecting Department of Commerce, Department of Education, Department of Health and Human Services, and the filters of "Exemptions" and "2015," the generated chart reflects that Exemption 6 is the most often cited reason for failing to release information).

[126] *See* Max Galka, *Transparent Censorship: An In-Depth Look at FOIA in 2015*, FOIA Mapper (April 11, 2016) *available at* https://foiamapper.com/annual-foia-reports-2015/ [hereinafter "FOIA Mapper"]. *See also* Arieff v. Dept. of Navy, 712 F.2d 1462 (D.C. Cir. 1983). Requesting parties may appeal denial decisions to the federal court, which reviews denials de novo and generally resolves FOIA disputes at the summary judgment stage. *See* Judicial Watch, Inc. v. Dep't of the Navy, 25 F. Supp. 3d 131, 136 (D.C. Cir. 2014). "If, however, the record leaves substantial doubt as to the sufficiency of the search, summary judgment for the agency is not proper." *See* Truitt v. Dep't of State, 897 F.2d 540, 542 (D.C. Cir. 1990). If the agency is able to provide responsive records, but a portion of the record should be properly withheld, the agency may not deny complete disclosure if the record can be segregated such that the exempt portions are redacted and the nonexempt portions are disclosed unless it is impossible to separate the two. 5 U.S.C. § 552(b) (2012).

the comparison chart below, the majority of declinations are based on privacy concerns, and under the Obama administration, those figures increased despite the executive branch's policy for transparent government.[127]



[128]

Exemption 6 is the leading exemption agencies use to avoid the release of "personnel and medical files and similar files the disclosure of which would constitute a 'clearly unwarranted invasion of personal privacy[.]'".[129] The Supreme Court determined that the phrase "similar files" within Exemption 6 includes data that, if released, would subject someone to "injury and embarrassment that can result from the unnecessary disclosure of personal data."[130] Only the balancing of the public interest in its right to know against a person's privacy interest in the particular data governs whether the data should be released, and not the type of file itself.[131] Further, exemptions, including those involving privacy, are to be narrowly construed in favor of release, and an agency must distinguish between a substantial and a de minimis privacy interest.[132] Examples of information that do not implicate privacy concerns include information not linked to a particular individual, and instances of federal employee information that is not personal in nature.[133] Despite these constraints, data privacy serves as an easy excuse for the agency to decline the request.

---

[127] *See* FOIA Mapper *supra* note 124.

[128] *Id*.

[129] *See* U.S. Dept. of State v. Washington Post Co., 456 U.S. 595, 602-03 (1982) (upholding the government's denial of requests for documents regarding whether particular Iranian nationals held valid United States passports).

[130] *Id*. at 599.

[131] *Id*.

[132] *See* Multi Ag. Media LLC v. Dept. of Ag., 515 F.3d 1224, 1230 (D.C. Cir. 2008) (noting "a privacy interest may be substantial-more than de minimis-and yet be insufficient to overcome the public interest in disclosure.").

[133] *See e.g.* Arieff v. US Dept of the Navy, 712 F.2d at 1467 (aggregate data). *See also* Aguirre v. SEC, 551 F. Supp. 2d 33, 54 (D.D.C. 2008) (employment termination data). The agency bears the burden of demonstrating that the exemption is appropriate. *See* Fed. Open Mkt. Comm. of Fed. Reserve Sys. v. Merrill, 443 U.S. 340, 352 (1979). *See* Exec. Office for US Attorney, 789 F.3d 204, 209 (D.C. Cir. 2015) ("An agency's task is not herculean[,] it must

Consider as an example a state law case--the Southern Illinoisan open record's request for information from the Illinois Department of Public Health about the incidence of neuroblastoma in Illinois from 1985 to the date of the request.[134] The agency denied the request because it determined the release would violate individual privacy in the database as researchers would be able to reverse engineer the released data with other publicly available data and identify the individuals by zip code and determine whether they had cancer.[135] On appeal, the Illinois Supreme Court found denying release of public data simply because someone could combine outside data with the released data to determine a person's identity was an inappropriate litmus test.[136] Without data from the Illinois Department of Public Health, and given the data restrictions imposed on hospitals and other health organizations by HIPAA, it would have been practically impossible to obtain this data from any other source.[137] The only practical solution would have been to drop the investigation altogether had the court not ordered the release.[138]

Rather than unnecessarily refusing to release data based on specific privacy concerns, agencies should be instructed by the federal court's decision in *Arieff v. Dept of the Navy*, finding the Navy's refusal to release general prescription data regarding 600 patients based on privacy concerns inappropriate.[139] Requesting parties asked the Navy for "all records concerning releases of any prescription drugs" between certain time periods.[140] The Navy denied the request in part because the release of the data "would constitute a 'clearly unwarranted invasion of [the] personal privacy' of the Beneficiaries in violation of Exemption 6."[141] The Court disagreed, noting any "secondary effect" of the release is irrelevant regarding whether the government FOIA release should occur.[142] Rather, the actual production of the documents must be the cause of the invasion of privacy for the withholding to be proper.[143] Unless it is apparent a person's identity will be revealed because of the agency's release of the data, rather than mere speculation that the release might identify an individual, the government should release the data. Failing to do so diminishes the sources of data for public researchers and increases the power of data brokers and the privatization of research, leading to data inequality and negative rent-seeking.

---

describe the justifications for nondisclosure with reasonably specific detail and demonstrate that the information withheld logically falls within the claimed exemption.") (internal quotations and citations omitted). A requesting party must not only show "more than a mere suspicion" that the agency acted negligently in denying disclosure and that the release will advance a significant public interest. *See* Nat'l Archives & Records Administration v. Favish, 541 U.S. 157, 172 (2004).

[134] *See* Southern Illinoisan v. The Illinois Dept. of Public Health, Civ. Doc. No. 98712, 844 N.E.2d 1 (Ill. 2006). Although a state law case, state open records laws are similar to FOIA and its exemptions.

[135] *Id*. at 3.

[136] *Id*. at 19.

[137] *Id*.

[138] The court made an interesting observation regarding this situation: "[t]he entire purpose of the Cancer Registry Act would be effectively repealed by subsection 4(d) if we did not impose the reasonableness requirement, because any fact, no matter how unrelated to identity can tend to lead to identity, and, therefore, any and every fact would be exempt under subsection 4(d)." *Id*. at 19.

[139] 712 F.2d 1462, 1465 (D.C. Cir.1983).

[140] *Id*. (internal quotations omitted).

[141] *Id*. at 1468.

[142] *Id*.

[143] *Id*.

B. Individual Protections under the Privacy Act

In contrast to FOIA, a statute of release, the Privacy Act is a statute of protection from release and is another hurdle for researchers to overcome.[144] The Privacy Act of 1974 governs "the collection, maintenance, use, and dissemination of information about individuals that is maintained in systems of records by federal agencies."[145] An individual may access and correct his own information contained within federal government's records.[146] Any information about the individual that is "linked to that individual by name or identifying particular[.]" is protected from government release.[147] Thus, where a FOIA exemption would permit the government to deny release of personal information, a requester may force its release only when the data pertains to *himself* rather than information regarding third parties.[148] Regarding a request for a third party's information, the Privacy Act prohibits federal agencies from disclosing personally identifiable information without the subject of the request's express consent.[149]

In those instances under FOIA that *require* the government to release information, the Privacy Act likewise *permits* the government to release the information.[150] However, Exemption 6 of FOIA indicates the government's release of personal information is purely discretionary and not mandatory; thus the Privacy Act would allow the government to withhold information falling within Exemption 6's parameters.[151] Accordingly, where there would be a "clearly unwarranted invasion of privacy" by the release of data under FOIA Exemption 6, the records may not be released without the subject individual's consent under the Privacy Act.[152] In this regard, agencies are left with unfettered discretion to release or not to release data.[153] The general practice is for agencies to decline to release records even where release might be possible.[154]

---

[144] 5 U.S.C. § 552 (2012).

[145] *Id*.

[146] *Id*.

[147] *See* Pierce v. DOJ, 512 F.3d 184, 188 (5th Cir. 2007), *cert. den.*, 553 U.S. 1019 (2008). 5 U.S.C. § 552a(a)(4) (2012). A record is "any item, collection or group of data about an individual that is maintained by an agency, including but not limited to his education, financial transactions, medical history, and criminal or employment history and contains his name, or the identifying number, symbol or other identifying particular assigned to the individual such as a finger or voice print or photograph." *Id*.

[148] 5 U.S.C. § 552a(d) (2012).

[149] *Id*. at § 552a(b) (2012).

[150] *Id*. at § 552a(b)(2) (2012).

[151] *See id*. at § 552(b)(6) (2012) and § 552a(b)(2) (2012).

[152] *Id*.

[153] A plaintiff's challenge to the government's release under the Privacy Act must demonstrate that "(1) the information is a record within a system of records, (2) the agency disclosed the information, (3) the disclosure adversely affected the plaintiff, and (4) the disclosure was willful or intentional." 5 U.S.C.A. § 552a (2012). *See* Luster v. Vilsack, 667 F.3d 1088, 1089 (10th Cir. 2011). For the government to be held liable, the release of data must have been "so patently egregious and unlawful that anyone undertaking the conduct should have known it unlawful." 5 U.S.C.A. § 552a(g)(4) (2012). *See* Maydak v. U.S., 630 F.3d 166, 179 (D.C. Cir. 2010). Third parties (nongovernmental officials) who release private data are not subject to the Privacy Act. 5 U.S.C.A. § 552a(b) (2012). *See also* N.L.R.B. v. Vista Del Sol Health Services, Inc., 40 F. Supp. 3d 1238, 1268 (C.D. Cal. 2014). Moreover, if the released data is available elsewhere rather than merely contained within the federal government's records, there is no Privacy Act violation if the government releases the same information. *See* York v. McHugh 850 F. Supp. 2d 305, 310-12 (D.C. Cir. 2012) and Doe v. US Dept. of Treasury 706 F. Supp. 2d 1 (D.C. Cir. 2009) (finding it only applies to releases of data directly or indirectly from a governmental system of records).

[154] Plaintiffs about whom data is sought may pursue a "reverse FOIA" claim seeking protection from the release of their data. *See* Doe v. Veneman, 380 F.3d 807, 810 (5th Cir. 2004).

The general practice makes researchers' use of free government data more difficult and contributes to the advent of data inequality.

Moreover, outside the context of FOIA, when, how, and if the government provides access to open data is the subject of administrative policy goals rather than specific laws.[155] Academics, legal scholars, lobbyists, and other stakeholders often try to influence the executive branch's policies.[156] Access to the White House enables individuals and interest groups potentially to influence existing and future policy.[157] The relevance of whether and how data giants have access to the White House dictates the likelihood the public and researchers will have access to free, unbiased, and fact-checked data rather than data for purchase or subject to trade secret protections. If future administrations allow commercial data brokers to dictate government policy, it will further the data inequality and weaken research credibility as researchers will turn to data brokers for their information. Only those researchers with adequate funding, or those who collaborate with data brokers, will contribute to the future of research.

C. Government's De-Identification of Data and Concomitant Privacy Concerns Unreasonably Dominate Its Release Decisions

We contend that generalized denials of access to government data and the government's anonymization techniques that do not consider a person's voluntary revelation of the information the government is trying to protect will promote further data inequality. Nonetheless, there are countervailing policy considerations associated with exposing one's personal data to the government: (1) concern over third parties' accessing private data through a FOIA request, or (2) concern over the efficacy of anonymization techniques used in the government's general release of statistics leading to revelation of private data. Many government services require personal data to utilize their service.[158] Initially, there may be a chilling effect and disincentive for individuals to provide the government with personal data, knowing the government may store it and that it may be subject to release.[159] The possibility of re-identification is significant particularly because data are no longer within the individual's control.[160] The data can be subject to misuse or abuse.[161] Because of these privacy concerns, individuals may be disinclined to inquire about personal services that are relevant to the public health sector like pregnancy, disease, drugs,

---

[155] *See generally, id.*

[156] One entity, Google, had the most significant contact with the Obama administration in small groups or individual meetings with key White House officials. Between January 2009 and October 31, 2015, Google met at the White House approximately 427 times. *See* Johnny Kampis, *Visitor Logs Show Google's Unrivaled White House Access* (May 16, 2016) available at http://watchdog.org/265252/visitor-logs-google-white-house. This exceeds the number of meetings all of the top 50 oil and gas companies had with the White House during the same time frame. *Id*. Because the White House is not subject to FOIA, whether the visitor logs actually capture all meetings is unclear. *Id*. Prior administrations did not make visitor logs publicly available, and there is no obligation that future administrations do so. *Id*. The Trump administration has indicated it will not release visitor logs. *See* Julie Hirschfeld Davis, *White House to Keep Its Visitor Logs Secret*, The New York Times (April 14, 2017) *available at* https://www.nytimes.com/2017/04/14/us/politics/visitor-log-white-house-trump.html?_r=1.

[157] *See* Altman et al. *supra* note 74 at 1993.

[158] *See* Borgesius et al. *supra* note 18 at 2088.

[159] *Id*.

[160] *Id*.

[161] *Id*. at 2088-92.

financial issues, or suicidal thoughts.[162] Although the chilling effect certainly can impact the individual, society as a whole is likewise impacted.[163] Privacy violations because of the government's data release have been few and far between.[164] Government agencies have done an admirable job of balancing the need for privacy while also providing the public with statistical data.[165] They have been at the forefront of developing tools and techniques to make this possible.[166]

For example, the Confidential Data Protection and Statistical Effective Act of 2002 ("CIPSEA") governs the government's release of statistical data.[167] CIPSEA's terms dictate how the federal government can prevent identification of an individual through the public release of statistics when aggregated from a variety of governmental sources involving that individual.[168] Agencies utilize techniques to prevent individuals' identification.[169] Based on privacy concerns, a variety of de-identification[170] tools have been espoused by experts and used by governmental agencies and are known as "statistical disclosure limitation techniques."[171] The purpose of statistical disclosure limitation techniques is to prevent the disclosure of an individual's identity or the attributes of particular individuals when data are released to the public in an aggregate form or released to researchers in the form of microdata.[172] These techniques include redacting personal identifiers, coarsening attributes such as modifying a person's location, recoding the values associated with a person into rounded values or intervals, swapping values in similar records, truncating extreme values, and adding random noise to the data.[173] These tools add background noise to the statistical data, making it more difficult accurately to identify a particular person in any aggregated materials.[174] These tools consider the impact the modifications have on the utility of the data as well as the extent to which they prevent unwarranted disclosure.[175] However, they

---

[162] *Id*. at 2090.
[163] *Id*.
[164] *See* Altman et al. *supra* note 74 at 1993.
[165] *Id*.
[166] *Id*.
[167] *See* Altman et al. *supra* note 74 at 1992.
[168] *Id*. at 1992.
[169] *Id*. at 2004. Many agencies have disclosure review boards or panels to ensure release does not breach privacy rights. *Id*. at 1994. A few months prior to the proposed release, the agencies compare their disclosure limitation techniques with the availability of similar data potentially linked to the proposed release data. *Id*. Other statistical disclosure laws may apply depending on the agency. *Id*.
[170] De-identification is a process or set of processes that utilize a variety of tools to mask and prevent the ability of a third party from identifying any one particular individual from an aggregated data set. *See* Simson L. Garfinkel, *De-Identification of Personal Data*, NISTIR 8053 (Oct. 2015) *available at* nv/pubs.nist.gov/nistpubs/ir/2015/NIST.IR.8053.pdf. Anonymization is also a method of preventing the identification of an individual in a data set and the identity of that individual remains unknown to the collector as well; whereas in de-identification, the identity of the individual may be known to the collector. *Id*.
[171] *See* Altman et *al*. *supra* note 74 at 1972. *See also* Ira S. Rubinstein, Woodrow Hartzog, *Anonymization and Risk*, 91 WASH. L. REV. 703, 712-713 (2016).
[172] *See* Jane Yakowitz, *Tragedy of the Data Commons*, 25 HARV. J.L. TECH. 1, 9 (2011).
[173] *See* Altman et al. *supra* note 74 at 1995.
[174] *Id*. at 1973.
[175] *Id.* The impact of de-identification is relevant because it decreases data's utility. *See* Rubinstein et al. *supra* note 169 at 709 (noting the failure of anonymization technology to protect privacy leading to polarization between policy makers) and Altman et al. *supra* note 74 at 1973-74.

do not take into consideration whether the subject has otherwise provided the information to a data broker. We contend this should be added as a consideration for analysis. If individuals freely provide information to data brokers, which the government likewise has possession of, the type of information provided and the type of data broker the information was provided to should be part of the government's assessment as to the sensitivity of the information. In those instances where the information is freely provided to a variety of online sources and the information does not relate to issues of identity theft or other sensitive information, the release might be appropriate with minimal statistical limitation techniques.

Further concern over the public's ability to combine discrete data sets in a manner that then identifies an individual has led the government to analyze the released data using a "mosaic effect."[176] In determining whether to release statistical or other mass data points, government agencies assess the impact the release will have on an individual's personal data and utilize a conservative approach to prevent disclosure of sensitive data.[177] The assessment is known as a "privacy first" assessment, discussed more fully in Section V(B) and, which we contend, will lead to the further privatization of research if not balanced with an assessment of voluntary revelation by the individual himself. It seems completely counter-productive to move the entire responsibility of protecting privacy to the government while allowing data brokers to operate completely unrestricted. If we are to prevent the government from releasing the type of data that the data brokers are free to sell, then the inevitable result is that the data brokers corner the market on data, resulting in another form of negative rent-seeking.

V. The Public's Ability to Access Data from Data Brokers and the Data Brokers' Privacy Obligations

Despite significant pressures on the government to ensure data privacy, a consistent regulation for data privacy gathered, distributed, or maintained by private entities has not emerged.[178] It is this unusual discrepancy between strong privacy obligations for the government and the nearly nonexistent privacy obligations for the data brokers that furthers a negative rent-seeking situation. In 1989, the Supreme Court stated: "the common law and the literal understandings of privacy encompass the individual's control of data concerning his or her person."[179] However,

[176] Defined by ComputerWorld as "data elements that in isolation look relatively innocuous can amount to a privacy breach when combined." *See* Jaikumar Vijayan, *Sidebar: The Mosaic Effect* (Mar. 15, 2004 12:00 a.m. PT) *available at* http://www.computerworld.com/article/2563635/security0/sidebar--the-mosaic-effect.html. Data experts agree that there is no "fool proof" way to ensure that disclosure limitation techniques will eliminate the ability of a third party to cull together data and identify a particular individual within a discrete data set. Both Netflix and America Online released anonymized data that others were able to compare with publicly available data to identify those members contained within their studies. *See* The Center for Open Data Enterprise, *Briefing Paper on Data and Privacy*, *available at* http://reports.opendataenterprise.org/BriefingPaperonOpenDataandPrivacy.pdf (last visited April 3, 2017).

[177] *Id*. *See* Altman et al. *supra* note 74 at 2001. Unlike public data sets, restricted data requires researchers to apply for access to the data, and the governmental release depends on a formal screening process. *Id*. at 1988. The use is limited to the purposes specified through data use agreements. *Id*. at 1998-2002. Regarding data related to individuals (financial, demographic, purchasing behavior, etc.), if the government agency refuses to release data, the researcher has the option of purchasing this data from data aggregators; likewise, data related to organizations have long been available from other sources (CRSP, COMPUSTAT, and others). *Id.*

[178] *See* Paul Ohm, *Sensitive Data*, 88 S. CAL. L. REV. 1125, 1140-41 (2015).

[179] *See* DOJ v. Reporters Committee for Freedom of Press, 489 US 749, 763 (1989).

the right to privacy outlined in the United States Constitution merely prevents the government from "intrusive government activities."[180] It does not protect individuals from private sector intrusion.[181] In the United States, courts treat personal data as a product rather than a right, as opposed to the European Union which considers these rights fundamental.[182] As such, a more balanced approach in favor of relaxing the government's privacy obligations along with an increase in access to data broker information is necessary.

A. Access to Public Data

In light of the governments' declination to release data, or to release only thoroughly scrubbed data, researchers have better access to purchased data.[183] Commercial entities have found data analytics to be big business.[184] Indeed, anyone can buy just about any data from a data broker.[185] Researchers have more access to privately gathered data, and commercial vendors of this type of research data are growing.[186] Commercial vendors include Datasift, Acxiom, Treato, and TrueLens.[187] Data brokers collect data from a variety of sources, including public records, loyalty cards, websites, social media, bankruptcy data, voting history, consumer purchase data, web browsing activities, and warranty registrations.[188] One of the nine (9) data brokers, Acxiom, studied by the Federal Trade Commission ("FTC"), has 3,000 discrete data segments for nearly every United States consumer.[189] Data brokers gather data and aggregate data into discrete categories, identifying consumers as *inter alia*, the "expectant parent," "bible lifestyle," and "financially challenged."[190]

Data brokers often gather data from the government, including through the Census Bureau's demographic studies (identifying "ethnicity, age, education level, household makeup, income, occupations, and commute times" along with "geographic data including roads, addresses,

---

[180] *See* Ashley Kuempel, *The Invisible Middlemen: A Critique and Call for Reform of the Data Broker Industry*, 36 NW. J. INT'L L. & BUS. 207, 214 (2016).
[181] *Id*.
[182] *Id*. at 215.
[183] Marketing of data gathered by data brokers' accounts for the largest amount of their revenue generation, followed by risk mitigation and people search products. *See* 2014 Data Broker Report *supra* note 10. Data brokers often require their clients to certify that they will not violate a federal law like the Federal Credit Reporting Act. *Id*. However, the data brokers do not monitor or review whether violations occur. *Id*. The only limitation on what can be purchased is the efficacy of the particular data broker. *Id*. Data brokers dictate the nature of the relationship with the consumer through standard contractual agreements. *Id*. Data brokers and their sources generally enter into one (1) of three (3) types of contractual relationships with purchasers: (1) outright ownership of the gathered data, (2) license to use the data for a certain time period, or (3) the right to resell the data. *Id*.
[184] *See* Kuempel *supra* note 178 at 209-210 and Breggin *supra* note 16.
[185] *Id*. at 210.
[186] *See* Michael Mattioli, *supra* note 25 at 558. When asked about their clients and data sources by a Senate Committee during the investigation, the brokers refused to reveal this data, but generically noted they work for "47 Fortune 100 clients", "5 of the 13 largest U.S. federal government agencies" and "both major national political parties." *See* Gregory Manus, *How Corporate Data Brokers Sell Your Life and Why You Should Be Concerned* (Aug. 24, 2015 2:27 p.m.) *available at* https://thestack.com/security/2015/08/24/how-corporate-data-brokers-sell-your-life-and-why-you-should-be-concerned/.
[187] *See* 2014 Data Broker Report *supra* note 10.
[188] *Id*.
[189] *See* Manus *supra* note 184.
[190] *Id*.

congressional districts, and boundaries for cities, counties, subdivisions, and school and voting districts"), the Social Security Administration's Death Master File (listing "consumer's names, social security numbers, and dates of death"), along with the U.S. Postal Service's address standardization and change of address data, and other data from federal and international agencies.[191] Additional governmental data is provided by state agencies, such as licensing records, real property records, taxes, voter registration, court records, and motor vehicle records.[192] The data brokers also filter social media and other Internet blogs and posts, garnering data when the user does not set privacy restrictions.[193] Finally, data brokers purchase data from commercial data sources like retailers learning a consumer's purchase histories, purchase prices, dates of purchase, and form of payment used, along with registration sites, such as news and travel sites.[194] Privacy concerns surrounding data brokers' gathering and selling consumer data is apparent, but regulation is limited.

With respect to these private entities, the origins of privacy law are based in tort or are contained within discrete sets of specific legislation, i.e. driver's license data, credit reports, etc.[195] The release of non-sensitive data does not necessarily result in harm, but others could use it to de-identify data that, when released, was released using anonymization tools to protect consumer privacy.[196] The main regulatory body protecting consumer privacy is the FTC which has authority over consumer data brokers under Section 5 of the Federal Trade Commission Act.[197] In 2014, the FTC issued a report regarding Data Brokers, entitled "Data Brokers A Call for

[191] *See* 2014 Data Broker Report *supra* note 10 (other agencies include the Federal Bureau of Investigation, U.S. Secret Service, and the European Union).
[192] *Id.*
[193] *Id*.
[194] *Id*.
[195] *See* Ohm, *supra* note 7 at 1732-34. In addition to the initiatives surrounding government records and their transparency, the Obama Administration attempted, unsuccessfully, to rectify some of the concerns over individual privacy rights and commercial data brokers through the Consumer Privacy Bill of Rights in 2012. *See* Natasha Singer, *Why a Push for Online Privacy Is Bogged Down in Washington*, (Feb. 28, 2016) *available at* https://www.nytimes.com/2016/02/29/technology/obamas-effort-on-consumer-privacy-falls-short-critics-say.html?_r=0. Likewise, the Data Broker Accountability and Transparency Act of 2015, and the Data Security and Breach Notification Act, which would have increased consumer rights, has repeatedly failed in committee. *See* Manus, *supra* note 184.
[196] *See* Amelia R. Montgomery, *Just What the Doctor Ordered: Protecting Privacy Without Impeding Development of Digital Pills*, 19 VAND. J. ENT. & TECH. L. 147, 157-158 (2016).
[197] *See* 15 U.S.C. § 45(c)(2) (2006). In 2016, the Federal Communications Commission ("FCC") attempted to join the field of privacy regulation for telephone and cable companies by enacting the Protecting the Privacy of Broadband and Other Telecommunications Services Order (the "Privacy Order"). *See* FCC Releases Rules to Protect Broadband Consumer Privacy (Nov. 2, 2016) FCC-16-148A1_Red.pdf *available at* https://www.fcc.gov/document/fcc-releases-rules-protect-broadband-consumer-privacy (last visited April 27, 2017). The Privacy Order was designed to limit the quantity of data a telephone or broadband provider collects about its consumers, including their "geo-location, financial data, health data, children's data, social security numbers, web browsing history, app usage history, and the content of their communications." *Id*. Consumers would have had to consent to the Internet service providers' use and sharing of such data for anything other than the purposes for which the broadband provider services the consumer, e.g. billing. *Id*. However, under the new administration these regulations have been revoked and are unlikely to be implemented any time soon. *See* Alina Selyukh, *As Congress Repeals Internet Privacy Rules, Putting Your Options in Perspective*, (Mar. 26, 2017 6:58 PM ET) (detailing ways consumers can protect their own privacy online and on smartphones).

Transparency and Accountability."[198] The FTC studied nine (9) data brokers, representing over 1,000 companies.[199] The Report addressed (1) marketing products, (2) risk mitigation products, and (3) people search products offered by data brokers.[200] The Report noted that consumers benefit from easier access to goods and services and to lower or free web services because these services derive financial benefits from consumers through the sale of specifically marketed advertisements based on the data brokers' data.[201] Nonetheless, the Report noted areas for improvement: (1) the need for transparency in data brokers' policies, (2) the "aggregation effect" leading to potentially discriminatory use of data, and (3) the potential security risks with stored data.[202]

Unfortunately, once the data broker or private entity sells the data to a third party, the FTC's jurisdiction likely ceases.[203] The FTC can regulate when a business sets a privacy policy or markets the privacy of its product, and the practice is found to be deceptive.[204] However, the FTC cannot require that a company set a privacy policy.[205] Those that do not have privacy policies and do not promise privacy to their customers are exempt from liability except for common law privacy tort claims.[206]

---

[198] *See* Kuempel *supra* note 178 at 234. *See also* 2014 Data Brokers Report *supra* note 10. (The FTC defined the term "data broker" as a company that "collect[s] consumers' personal data and resell[s] or share[s] that data with others.").

[199] *See* 2014 Data Brokers Report *supra* note 10 (the nine (9) data brokers who received FTC requests for data were: Acxiom, Corelogic, Dataogix, eBureau, ID Analystics, Intelius, PeekYou, Rapleaf, and Recorded Future).

[200] *Id*.

[201] *Id*.

[202] *See* Kuempel *supra* note 178 at 218. *See also* 2014 Data Brokers Report *supra* note 10. The FTC recommends four (4) basic areas in need of legislative action, including (1) requiring data brokers to provide consumers with access to their data, including sensitive data to a reasonable level of detail, (2) allowing consumers the option of opting-out of having the data shared for marketing purposes, (3) informing consumers of the source of their data so that they can correct any inaccurate data, and (4) obtaining a consumer's prior affirmative consent where sensitive data is being collected. *Id*.

[203]*See* Kwame N. Akosah, *Cracking the One Way Mirror: How Computational Politics Harms Voter Privacy and Proposed Regulatory Solutions*, 25 FORDHAM INTELL. PROP. MEDIA & ENT. L. J. 1007, 1046 (2015). Other privacy laws may assist in regulating big data's use of certain data. *Id*. *See* Eric Everson, *Privacy by Design: Taking Control of Big Data*, 65 CLEV. ST. L. REV. 27, 37 (2016) (noting the wide array of federal and state laws targeting privacy issues).

[204] *See* Lipman *supra* note 112 at 790-92.

[205] *Id*.

[206] *Id*. However, the FTC has had some success in this area with one administrative action against Google for its misrepresentation of what it collected when consumers utilized Apple's Safari Internet. *See* Victoria D. Baranetsky, *Social Media and the Internet: A Story of Privatization*, 35 PACE L. REV. 304, 331 (2014). In another FTC administrative compliance case, the FTC found Facebook deceived its users by allowing their data, concerning those items they marked as "like", to be public even though Facebook lead its users to believe it was private. *Id*. at 332. Whether the Defend Trade Secrets Act of 2016, an amendment to the Economic Espionage Act, will provide any additional privacy rights to individuals is yet to be seen. 18 U.S.C. § 1839(3) (2016) (providing individuals with a federal private cause of action where the plaintiff has "taken all reasonable steps necessary" to keep their data secret and the data "derives an independent economic value" if made generally known to the public). This statute could form the basis of a private action against data brokers' and their repurposing of consumer data without their knowledge.

Interestingly, data brokers are somewhat regulated by their industry trade associations that have identified best practices for handling consumer data.[207] Although self-regulation provides some guidance for protecting consumer data, it is purely voluntary.[208] Companies' policies concerning access to customer data are inconsistent.[209] A joint task force made up of various governmental and nongovernmental stakeholders, including web application providers, has suggested certain best practices.[210] One suggestion is that providers issue a "data disclosure chart" which would require applications to display the types of data their application collects from the user.[211] Because these best practices are purely voluntary, not all entities implement them.[212] Thus, it is unlikely self-regulation will lead to any meaningful privacy or release practices in this area without some governmental or industry incentives.

B. Conflicting Theories: Governmental Release vs. Private Entity Release of Aggregated Data

In situations where the government agency refuses to release the data for privacy reasons, and denies the researcher access to public data, the researcher has few alternatives, but may use surrogate measures, or alter the research question, or drop the inquiry altogether.[213] Research outcomes are adversely affected by all three (3) options.[214]

Historically researchers attempted to strike a balance between the risk of disclosing personal data against the usefulness of the released data when requesting government data, which used statistical disclosure techniques discussed in Section IV(C) above.[215] About ten (10) years ago, researchers from Microsoft developed what is known as the "privacy first" or "differential privacy" model, meaning a person's privacy interest is more significant in determining whether to release data than any other consideration.[216] Differential privacy, coined by Cynthia Dwork a highly respected computer scientist and researcher with Microsoft, is a procedure for assessing the risk resulting from the data release, whether through government release or otherwise.[217] These researchers claim data release should be analogous to encryption considerations.[218] Differential privacy can be summarized as follows: how do we release data from a dataset consisting of n records so that a malicious user who has access to the true values of (n – 1) of

---

[207] *See* Kuempel *supra* note 178 at 216. *See also* Singer supra note 193 (discussing how the industry's self-regulation was designed to work in conjunction with the Department of Commerce). However, lack of consensus has inhibited solid industry self-regulation.
[208] *See* Kuempel *supra* note 178 at 216-217.
[209] *Id*. at 217.
[210] *Id*.
[211] *See* Singer *supra* note 193.
[212] *Id.* Other forms of best practices in data gathering of face recognition and voice recognition technologies have failed, stalling any cooperative self-regulatory efforts. *Id.*
[213] *See* Rubinstein *supra* note 169 at 719-720
[214] Freedom of Information Act, 5 U.S.C. § 552 (2006), as amended by OPEN Government Act of 2007, 5 U.S.C. § 552 (Supp. II 2008). *See* Rubinstein *supra* note 169 at 754.
[215] *See* Altman *supra* note 74 at 1977.
[216] *See* Chin et al., *supra* note 98 at 1427 (2012). *See also* Facebook,
"*Data Use Policy*", *available at* http://www.facebook.com/full_data_use_ policy (last updated Sept. 23, 2011).
[217] *See* Chin *supra* note 98 at n. 17.
[218] *Id*. at 1422 & notes 17, 42 (citing opposing legal scholars).

those records would not be able to infer data about the remaining nth record.[219] These researchers contend that the data available from data brokers makes this a likely scenario, i.e. the malicious user can purchase data about the (n – 1).[220]

One prime example the privacy first theorists point to is the inadvertent sharing of anonymized data from subscribers of Netflix.[221] As part of a marketing contest, Netflix allowed anyone to register for a chance to win $1 million for creating a movie rating system that was better than its existing system and would provide contestants with data on "training data set consist[ing] of more than 100 million ratings from over 480 thousand randomly-chosen, anonymous customers on nearly 18 thousand movie titles."[222] Thereafter, a follow-up contest provided access to additional customer data including zip codes, ages, gender, genre ratings and previously chosen movies.[223] A group of researchers accessed the data by registering for the contest; and, instead of modifying the formula, the researchers reverse engineered the data identifying the Netflix customers by comparing publicly available records (despite Netflix's de-identification of customer data through the assignation of random identifiers and deliberately "perturbing" the data by "deleting ratings, inserting alternative ratings and dates, and modifying rating dates.").[224] The researchers demonstrated that if someone knows a bit about a person and their viewing habits, say for example an office colleague's discussions at work, they could take that data in combination with the data provided by Netflix and determine that particular person's viewing habits.[225] Plaintiffs sued Netflix for this breach of their personal privacy and reached a settlement with Netflix. [226] At its core, Netflix was attempting to release useful data that could be used to further its goal of improving services, and yet did so exposing users' privacy interests even though it anonymized the data sets. [227] Clever individuals were able to use the aggregate data in combination with other sources to de-identify some individuals despite the anonymization technology.[228]

Differential privacy theorists recommend that the public release of data be modified to account for this possibility regardless of what the data reveal (sensitive or otherwise).[229] Certainly, privacy is an important consideration; however, a default rule of privacy first may not be the best method to protect data. Researchers have shown that in some cases differential privacy may result in meaningless data.[230] Also, one of the primary concepts of differential privacy is that all aspects of the protection mechanism must be transparent.[231] Practical implementations of

---

[219] *Id*. at 1422.
[220] *Id*.
[221] *See* Wu, *supra* note 113.
[222] *Id*.
[223] *Id*.
[224] *Id*.
[225] *Id.* at 1120.
[226] *Id*.
[227] *Id*. at 1121.
[228] *Id*.
[229] *Id*.
[230] *See* Jane Bambauer, Krishnamurty Muralidhar, Rathindra Sarathy, *Fool's Gold: An Illustrated Critique of Differential Privay*, 16 VAND. J. ENT. & TECH. L., 701-755 (2014).
[231] *See* Cynthia Dwork, *Firm Foundation for Private Data Analysis*, 54 COMMS OF THE ACM, 86, 91 (2011) *available at* http://research.microsoft.com/pubs/l16123/dwork_cacm.pdf (last visited April 27, 2017).

differential privacy have been anything but transparent.[232] For example, Facebook apparently utilizes a form of "differential privacy" in its advertisement targeting databases which allow an advertiser to target specific users.[233] However, the details of the methods used are entirely unknown to the general public.[234] Moreover, Apple recently announced it would implement a "differential privacy" style process without explanation of its methods.[235] Commentators criticized the failure to divulge how the "differential privacy" techniques are utilized, noting "[i]n the end, one must compare the reduction in harm actually afforded by using differential privacy with the increase in harm afforded by corporations having another means of whitewash, and policy-makers believing, quite wrongly, that there is some sort of cryptomagic to protect people from data misuse."[236]

This lack of transparency from Facebook and Apple is not surprising. Recently, advertisers (such as AT&T and Johnson & Johnson) pulled their advertisements from Youtube and Google because they found that their advertisements were appearing on websites that promote hate.[237] One would think that it would be easy for this technology giant to fix the problem instantaneously. The algorithms used to place the advertisements are so complex that they have not been able to assure the advertisers that their advertisements will not appear on inappropriate websites.[238] Astonishingly, Google itself seems to have been unaware of this problem.[239] The use of complex algorithms whose inner workings we cannot easily comprehend is a simple way for technology giants to build a smoke screen to protect their operations from the public.[240] The use of opaque differential privacy is consistent with this approach as it lulls the public into feeling secure and not protest the data gathering.[241] Differential privacy becomes the panacea to cure all data disclosure ailments when it is, in fact, a placebo.

It is also interesting that the entire responsibility of protecting the privacy of data is transferred to the public sector while data brokers are free to sell any and all data. The process of disclosure is fraught with uncertainty since the malicious user can never be certain about the identity of the individual or the values of the variables for that record.[242] Purchasing data from the data broker is a better option for the malicious user if we consider accuracy, effort, and cost; yet, there is no

---

[232] *See* Andy Greenberg, *Apple's 'Differential Privacy' Is About Collecting Your Data But Not Your Data* (June 13, 2016 7:02 p.m.) *available at* https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/. *See also* Matthew Green, *What Is Differential Privacy* (June 15, 2016) *available at* https://blog.cryptographyengineering.com/2016/06/15/what-is-differential-privacy/.

[233] *See generally*, Chin *supra* note 1 at 98.

[234] *Id*. at 1452-54.

[235] *See* Greenberg, *supra* note 230. *See also* Green, *supra* note 230.

[236] *See* Phillip Rogway, *The Moral Character of Cryptographic Work* (Dec. 2015) available at http://web.cs.ucdavis.edu/~rogway/papers/moral-fn.pdf.

[237] *See* Nick Statt, *AT&T and Verizon pull ads from Google and YouTube over hate speech* (Mar. 23, 2017 6:09 PM EDT) *available at* http://www.theverge.com/2017/3/22/15029214/att-verizon-google-youtube-pull-boycott-hate-speech.

[238] *Id*.

[239] *Id*.

[240] *Id*.

[241] *Id*.

[242] *See* Ohm, *supra* note 7 at 1710-1711 (discussing the various proponents for anonymization of data prior to governmental release and the governments' apparent acceptance of anonymization as the panacea for public release of data).

protection against the sale of this data.[243] For all we know, the person who could be identified through reverse engineering has already identified himself in some other consensual manner, or, data about that individual are already in the public domain. [244] Use of the "privacy first" assessment is guaranteed to prevent easy access to government data and yet not protect against a data broker's release of sensitive information.[245] The double standard contributes to data inequality and further negative rent-seeking, i.e. making access to government data more difficult, rendering the data held by the data brokers and the technology giants more valuable.

In an attempt to equal the playing field, many commentators and scholars argue that data brokers, like the federal government, should be regulated by best practices known as "fair information principles" ("FIPs").[246] All privacy statutes incorporate the best practices for computer databases to ensure that a person who provides data for one purpose is not subjected to the use of that data for other purposes without his prior consent.[247] FIPs are a set of best practices for the collection, storage, and use of personal data by the government, and the private sector.[248] The underlying philosophy of FIPs was the impetus behind the enactment of the Privacy Act discussed above in Section IV(B).[249] However, scholars disagree as to whether, and to what extent, a mandatory application of FIPs for data brokers' activities would infringe upon their First Amendment rights.[250] A court examines the nature of the speech to determine whether a FIP or any other regulation on the distribution or receipt of data implicates First Amendment concerns.[251] Where data brokers direct communications to consumers encouraging them to buy more services or products, this form of speech is commercial in nature and subject to the *Central Hudson* four-part analysis: (1) is the activity lawful and not misleading, (2) if so, then government may only restrict speech if "(1) it has a substantial state interest in regulating the speech, (2) the regulation directly and materially advances that interest, and (3) the regulation is no more extensive than necessary to serve the interest."[252] Arguably, the collection and use of data by one commercial entity which resells it to a third party with the ability of the customer to opt-out of such reuse promotes both a substantial state interest and is narrowly tailored to this interest.[253] Moreover, requiring the data broker to make the underlying nature of its data

[243] Ohm also challenges the theory that anonymization of data actually protects individual privacy. *Id*. at 1732.
[244] *Id*.
[245] *Id.*
[246] *Id.*
[247] *See* Borgesius et al. *supra* note 18 at 2101.
[248] *See* Richards, *supra* note 79 at 1513.
[249] *Id*.
[250] On one end of the spectrum, Eugen Volokh contends that most data privacy rules violate free speech, i.e. "violate my right to speak about you." *See* Eugene Volokh, *Freedom of Speech and Data Privacy: The Troubling Implications of a Right to Stop People from Speaking About You*, 52 STAN. L. REV. 1049, 1115-17 (2000). On the other end of the spectrum, Neil Richards advocates that FIPs "do not restrict the flow of data" but rather should be construed as confidentiality tools and that "data" are not entirely equal to "speech." *See* Richards, *supra* note 79 at 1533.
[251] *See* U.S. West, Inc. v. F.C.C., 182 F.3d 1224, 1232 (10th Cir. 1999), *cert. den.*, 530 U.S. 1213 (2000).
[252] *See* Central Hudson Gas & Elec. Corp. v. Public Serv. Comm'n of N.Y., 447 U.S. 557, 564-66 (1980). *See also*, U.S. West, Inc. v. FCC, 182 F.3d at 1235 (noting that while advancement of a privacy interest may be substantial, the government must articulate specifically how the regulation advances that interest, e.g. to avoid ridicule or harassment, rather than generically stating the restrictions are designed to protect privacy).
[253] *See* U.S. West, Inc. v. FCC, 182 F.3d at 1239 (noting opt-out strategies from solicitations are less restrictive alternatives). Subsequent to the court's ruling against the FCC's opt-in regulations, the FCC modified them to apply in instances where customer's data are distributed to third parties rather than to the customer's carrier alone. *See*

available for corroboration promotes a substantial governmental interest in decreasing negative rent-seeking.

VI. Opaqueness of Data Brokers' Data and Research Results

As previously noted, private entities have little restriction on where they get their data, or how they share their data.[254] Based on intellectual property and trade secret protections, the release of their data is opaque, and yet, there are no consequences to this lack of transparency or lack of data protection.[255] On the other hand, the government must protect the public data in its possession and must be transparent about its protection.[256] The fundamental discrimination between public and private sources of data, if not addressed, will lead to negative data inequality.

Due to a fundamental lack of transparency, it is unclear when, how, and what data are gathered by data brokers.[257] Data brokers gather much of the data without the specific knowledge or consent of the consumer.[258] Disconcertedly, a small number of sites receive the largest amount of traffic, meaning certain data aggregators and news sources control the majority of data consumers receive.[259] From a commercial perspective, approximately 81% of consumers conduct online research before purchase.[260] Forty-four percent of consumers commence their product search on Amazon's website, followed by Google, Bing, and Yahoo.[261] Regarding all searches, commercial or otherwise, website users search Google 100 billion times in a month.[262] Not surprisingly, a top priority for marketers is how to improve their Internet presence, and 72% of marketers found that ensuring their content was relevant to a consumer has been the most effective tool for their business.[263] With the consumer visiting various sites, even though the

---

Consumer Proprietary Network Data, 72 Fed. Reg. 31948 (June 8, 2007) (final rule codified at scattered sections of 47 C.F.R. pt. 64) and Telecommunications Carriers' Use of Customer Proprietary Network Data and other Customer Data, Third Report and Order and Third Further Notice of Proposed Rulemaking, 17 F.C.C. Rec. 14860, ¶32, 50, and 64. Additionally, proponents of restricting the reuse of data gathered from data brokers have supported a bill known as *Reclaim Your Name available at* https://www.ftc.gov/sites/default/files/documents/public-statements/reclaim-your-name/130626computersfreedom.pdf (last visited April 3, 2017).

[254] *See* Mattioli *supra* note 25 at 544-45.

[255] *Id*.

[256] *Id*.

[257] *See* 2014 Data Brokers Report *supra* note 10.

[258] *Id*. Certainly, Internet users bear some responsibility to manage the data they share online and this concept is contained within the Fair Data Practice Principles ("FIPPs") (FIPs and FIPPs are used interchangeably throughout). *See* Solove, *supra* note 64 at 1882. Initially, these principles were designed to address part of the government's concern over the increase in digital data and including "(1) transparency of record systems of personal data, (2) the right to notice about such record systems, (3) the right to prevent personal data from being used for new purposes without consent, (4) the right to correct or amend one's records, and (5) responsibilities on the holders of data to prevent its misuse." *Id*. The underlying theme of FIPPs is a user's awareness that data are gathered and that the user consents to the gathering of the data. *Id*.

[259] *See e.g.* Amy Mitchell, Jesse Holcomb, *State of the News Media 2016*, PEW RESEARCH CENTER (June 15, 2016) *available at* www.stateofthemedia.org/2010/special-reports-economic-attitudes/nieslsen-analysis.

[260] *See* Hubspot, *The Ultimate List of Marketing Statistics*, *available at* https://www.hubspot.com/marketing-statistics (last visited April 3, 2017).

[261] *Id*.

[262] *Id*.

[263] Video is increasingly a more popular tool for marketers through Youtube and Facebook videos. *Id*.

user reveals limited data on each, data can be aggregated by data brokers and compiled into a more detailed picture of the person and his or her private data.[264]

Many legal scholars criticize the inability to assess big data's pedigree as interfering with others' reuse of the research.[265] Traditional research methods define the research question, gather the data from a relevant data set, form a hypothesis, and test the hypothesis.[266] Once researchers publish traditional research, others test and challenge the research.[267] Modern commercial research alters this traditional method because big data often are considered proprietary in nature and not openly accessible for further analysis.[268] Accordingly, transparency of data is paramount to ensure accurate and thorough public research. Advancements in a data broker's techniques and algorithms are leading to more specific data, which are beneficial for marketing purposes but expose individuals to de-identification without their knowledge. For example, consumers are unaware that a grocery store can sell their purchasing data to third parties, and these third parties can then market to that consumer based on their grocery store purchase.[269] Moreover, website trackers can de-anonymize web browsing by linking to a person's Twitter and other social media accounts based on the person's clicking on the website link contained within the particular social media site.[270] In this regard, a person is now identified through the registration of their social media account and tied to the link, which is an indicator of interest in the content provided.[271]

Ideally, legislators could resolve this dilemma by requiring data brokers provide access to their underlying data used in research. In those instances where the data broker does not wish to divulge the underlying data, they should be required to include a disclaimer noting the data is protected by trade secrets and not subject to independent review. As it is unlikely any such legislation would be enacted by the current administration, government or industry incentives should be considered as discussed below.

A. Opaque Data Can Lead to Erroneous Interpretations

---

[264] *See* Solove *supra* note 64 at 1888-1889.
[265] *See* Mattioli *supra* note 25 at 544-45. *See also* 2014 Data Brokers Report *supra* note 10.
[266] *See* Omer Teno, Jules Polonetsky, *Judged by the Tinman: Individual Rights in the Age of Big Data*, 11 J. TELECOMM. & HIGH TECH. L. 351, 354 (2013). *See also* Eszter Hargittai, *Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites*, 659 ANNALS AM. ACAD. POL. & SOC. SCI. 63, 73 (2015) (identifying common issues with the use of certain social media sites to conduct studies and noting females tend to use Twitter and Tumblr the most, while less economically privileged do not; African Americans use Twitter while Asian Americans were less likely to use LinkedIn). *See also* Andrew Morarcsik, *Tranparency: The Revolution in Qualitative Research*, *available at* http://www.princeton.edu/~amoravcs/library/transparency.pdf (last visited April 3, 2017) ("Transparency is the cornerstone of social science. Academic discourse rests on the obligation of scholars to reveal to their colleagues the data, theory, and methodology on which their conclusions rest. Unless other scholars can examine evidence, parse the analysis, and understand the processes by which evidence and theories were chosen, why should they trust-and thus expend the time and effort to scrutinize, critique, debate, or extend-existing research.").
[267] *See* Teno *supra* note 264 at 355.
[268] *Id*. *See also* Lev Manovich, *Trending: The Promises and the Challenges of Big Social Data* http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data (2011).
[269] *See* Kuempel *supra* note 178 at 219.
[270] *See* Craig Mehall, *Study Finds Anonymous Browsing History Linkable to Individuals*, CQ Roll Call 2017 WL 370246 (Jan. 26, 2017).
[271] *Id*.

In addition to the inability to challenge research based on data purchased from data brokers, use of the data can lead to erroneous conclusions. Research suggests there is a potential for incorporating errors and biases at every stage of the data and research process.[272] According to the FTC's 2014 Data Broker's Report, some data brokers check the reliability of their data to ensure data are "internally consistent, corroborated by other sources, verifiable as legitimate, and that it encompasses a sufficiently large portion of the population."[273] However, data brokers do not assess the accuracy of the government or other publicly available data that they gather and then incorporate into their analysis.[274]

In this regard, the choice of the dataset used to make predictions, defining the problem to be addressed through big data, and the decisions based on the results of big data analysis could lead to potential discriminatory harms and are examined through examples below: (1) the advent of fake news and the public's belief in such news, (2) the effect of inaccurate background checks on employees and others, and (3) the unintended consequences of misinterpreting data.[275] Other researchers have noted that these concerns are overstated or are simply not new, and emphasize that rather than disadvantaging minorities, big data can create opportunities for low-income and underserved populations because the data can identify discrepancies or previously unknown needs.[276] However, it is becoming increasingly apparent that data can be manipulated either intentionally or unintentionally.[277] One thing that both public and nonpublic data has in common is the effect human judgment can have on the accumulation and assessment of the data.[278] The outcome of the analysis is dictated by the type data collected, the question presented, the pool of subjects in the data set, the method of collection, and its assessment. The method of culling and trimming data is known as "cleaning" the data.[279] The process is highly subjective, and the same data analysis could lead to different results depending on the person(s) conducting the analysis.[280] Because of this highly subjective method of research, proponents for data transparency in research are growing.[281]

1. Fake News

The most significant example of the need to ensure data's accuracy can be seen in the aftermath of the 2016 elections and the idea of fake news. Interestingly, people tend to believe what they

---

[272] For example, "social sorting involved obtain[ing] personalized group data in order to classify people and population according to varying criteria, to determine who should be targeted for special treatment, suspicion, eligibility, inclusion, access and so on." *See* Borgesius et al. *supra* note 18 at 2092.
[273] *See* 2014 Data Brokers Report *supra* note 10.
[274] *Id*.
[275] *See* Tene *supra* note 69 at 353.
[276] *Id*. at 355. *See also* Edith Ramirez et al, FTC REPORT: *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues* (Jan. 2016) *available at* https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf (detailing data gathered through public workshops regarding big data).
[277] *See* Tene *supra* note 69 at 355-57.
[278] *See* Mattioli *supra* note 25 at 559.
[279] *Id*. at 561.
[280] *Id*.
[281] *Id*.

read.[282] The fake news that dominated Facebook preceding the election was created by teenagers in Macedonia to make a profit from the pro-Trump movement.[283] The fake news stories were the impetus behind a gunman's attempt to kill the owner of a Washington, DC pizzeria under the mistaken belief, based on fake news circulating on Facebook, that Hillary Clinton and the restaurant owner ran a child sex ring out of the restaurant.[284] Even though the story appeared outlandish, people including the perpetrator believed the fake news.[285]

Despite the numerous discussions of "fake news," 39% of the American population are very confident that they can spot fake news, and 45% feel somewhat confident.[286] However, an Ipsos poll conducted for Buzzfeed News found 75% of Americans believed the fake news stories they had heard from the election.[287] During the election, Facebook had altered its algorithms that prioritized what its users saw by decreasing news media feeds and increasing posts and updates from friends and families.[288] Because Facebook owns the data, only it can say whether this helped the proliferation of fake news and whether their current efforts have reduced the circulation of fake news.[289] 1.8 billion monthly users and nearly half of all American adults read Facebook news.[290] Facebook has agreed to partner with third parties to flag fake news articles and alert users before they share the false news.[291] Google and Facebook collaborated with journalists to curb fake news stories in advance of the French elections.[292] Their efforts were carried out through trending and data mining techniques to detect problematic stories and provide crosschecking resources for readers and attach warning labels to suspect stories.[293]

---

[282] *See* Dave Davies, *Fake News Expert on How False Stories Spread and Why People Believe Them*, Radio Broadcast Transcript *available at* www.npr.org/2016/12/14/505547295/fake-news-expert-on-how-false-stories-spread-and-why-people-believe-them (last visited April 3, 2017) (interview with Craig Silverman of BuzzFeed News who has spent years studying media inaccuracy).

[283] *Id*.

[284] *See* Brett Edkins *Americans Believe they Can Detect Fake News-Studies Show They Can't* (Dec. 20, 2016 1:46 PM) *available at* www.forbes.com/sites/brettedkins/2016/12/20/americans-believe-they-can-detect-fake-news-studies-show-the-cant/#3f6796e54a4f.

[285] *Id*.

[286] *See* Michael Barthel, Amy Mitchell, Jesse Holcomb, *Many Americans Believe Fake News Is Sowing Confusion*, Pew Research Center, (Dec. 15, 2016) *available at* www.journalism.org/2016/.

[287] *Id*. (noting that 84% believed the fake news story that "Donald Trump Sent His Own Plane to Transport 200 Stranded Marines" and ¾ of Trump supporters believed the news story that Pope Francis endorsed Donald Trump). A Stanford University study supports the findings that Americans rarely identify fake stories as false. *See* Krysten Crawford, *Stanford Study Examines Fake News and the 2016 Presidential Election* (Jan. 18, 2017) *available at* news.stanford.edu.

[288] *See* Jane R. Bambauer, *All Life Is An Experiment (Sometimes It Is a Controlled Experiment)*, 47 LOY. U. CHI. L. J. 487, 499 (2015).

[289] *Id*. at 504-05.

[290] *Id*. The New York Times calls the republishing of fake news a "digital virus." The Editorial Board, *Facebook and the Digital Virus Called Fake News* (Nov. 19. 2016) *available at* https://www.nytimes.com/2016/.

[291] *See* Ivana Kottasova, *Google and Facebook are partnering with journalists to help prevent fake news stories from spreading during France's presidential election*, (Feb. 6, 2017 9:40 AM ET) *available at* money.cnn.com/2017/02/06/technology/france-elections-fake-news-facebook-google. Facebook will post "Disputed by 3rd Party Fact-Checkers" beneath the fake stories, will send an alert when the story is shared, and will rank the disputed stories lower in the news feed and will prevent these fake stories from transforming into promotions. *Id*. Likewise, Google intends to prohibit websites from selling fake news. *Id*.

[292] *Id*.

[293] *Id*.

While fake news is not a privacy issue, it has one important lesson: in the absence of transparent data, research will be unreliable, and researchers will not be incentivized to provide quality or accurate research as others will not be able to easily cross-check their research.[294] Researchers and those funding the research are responsible for ensuring transparency.[295] For example, the federal government often requires federally-funded research to be made public within a certain timeframe, journal editors require adherence to certain publishing guidelines, and some provide meaningful consequences for research misconduct.[296] These rules do not apply to nonfederally-funded research. In addition to the obvious potential inaccuracy issues, the use of big data poses potential ethical problems for society.[297]

2. Inaccurate Information and Credit Reports

According to one of the leading and largest data brokers, 30% of data brokers' data are inaccurate.[298] One expensive example was detected by the FTC when Spokeo, a data broker, marketed an employment screening tool with inaccurate profiles.[299] In addition to the $800,000 fine, one affected consumer sued Spokeo under the Fair Credit Reporting Act ("FCRA") for the publication of inaccurate data about his personal and employment background, believing it will harm his future employment possibilities.[300] In May 2016, the United State Supreme Court remanded the matter to the Ninth Circuit Court of Appeals to determine whether the consumer had alleged a concrete and particularized injury from Spokeo's violation of the FCRA but

---

[294] *See* Andrew Moravcsik, *Transparency: The Revolution in Qualitative Research* (2017) *available at* http://www.princeton.edu/~amoravcs/library/transparency.pdf. According to the American Political Science Association, transparency in research has three distinct characteristics: (1) data transparency-providing the reader with the evidence used to support the claims, (2) analytic transparency-"the process by which an author infers that evidence supports a specific descriptive, interpretive, or causal claim" and (3) production transparency-provides the reader with the facts surrounding the reason the author chose a particular source for his research. *Id*.

[295] *See* Patricia K. Baskin, *Transparency in Research and Reporting: Expanding the Effort through New Tools for Authors & Editors* (July 20, 2015) *available at* http://www.editage.com/insights/transparency-in-research-and-reporting-expanding-the-effort-through-new-tools-for-authors-and-editors.

[296] *Id*.

[297] *See* Danah Boyd, Jacob Metcalf, *Example "Big Data" Research Controversies, Data & Society Research Institute, Draft Version*, (Nov. 10, 2014) *available at* http://bdes.datasociety.net/wp-content/uploads/2016/10/ExampleControversies.pdf (identifying the ethical concerns between data supplied for certain surveys or commercial purposes being re-tooled and utilized for other purposes, i.e. governmental policy decisions). *See also*, Tene *supra* note 69 at 354 (citing cancer research example where data collector had some patients identifying themselves as gender unknown and analyzing which gender by height and weight).

[298] *See* Lipman, *supra* note 112 at 782. *See also* Bobby Allyn, "How the careless errors of credit reporting agencies are ruining people's lives" The Washington Post (September 8, 2016) available at https://www.washingtonpost.com/posteverything/wp/2016/09/08/how-the-careless-errors-of-credit-reporting-agencies-are-ruining-peoples-lives/?utm_term=.a8083de9383b (last accessed June 27, 2017). *See also* "Civil Rights, Big Data, and Our Algorithmic Future, A September 2014 Report on Social Justice and Technology" available at https://bigdata.fairness.io/ (last visited June 27, 2017) (detailing racial bias issues in the government run work-eligibility of potential employees which carries a 20% higher error rate for those who are foreign born as opposed to those born in the United States).

[299] *See* Press Release, Fed. Trade Comm'n, *Spokeo to Pay $800,000 to Settle FTC Charges Company Allegedly Marketed Data to Employers and Recruiters in Violation of FRCA* (June 12, 2012) available at https://www.ftc.gov/news-events/press-releases/2012/06/spokeo-pay-800000-settle-ftc-charges-company-allegedly-marketed [hereinafter Press Release, Spokeo to Pay $800,000].

[300] *Id*.

implicitly recognized there might be instances where such an injury automatically exists.[301] The FCRA applies to "consumer reporting agencies" that compile date into "credit reports" which are then used to score an individual's creditworthiness.[302] Credit reports are reports that contain "any data …bearing on a customer's credit worthiness, credit standing, credit capacity, character, general reputation, personal characteristics, or mode of living."[303] Data must be "used or expected to be used or collected" to serve as "a factor in establishing the consumer's eligibility for credit, insurance and employment for it to be subject to the FCRA.[304] If data are subject to the FCRA, the reporting agency must comply with a variety of obligations surrounding the collection, use, and right to challenge the data.[305] The trouble with this in the context of big data aggregation is that data mining can often be inaccurate and provide skewed research results, culminating in a lenders' or employer's decision to decline to loan to, or hire, a person.[306] Data brokers must be vigilant in the use of the data they gather and sell to avoid intentional discrimination claims. More importantly, access to transparent data is paramount so that researchers can ferret out instances of intentional discrimination. We assert the continued failures to do so will lead to data inequality and negative rent seeking.

3. Misinterpretation of Data

Biases can occur at any stage in this modern form of research, including the collection and the analysis stages.[307] If one considers the assessment of big data once the anonymization process is employed, one can see how easily research results could be faulty based on the important need for individual privacy.[308] Researchers identified an exemplification of an unintentional erroneous analysis after Hurricane Sandy and the 20 million tweets and data from FourSquare between October 27th and November 1st.[309] Some of this data reflected grocery purchases that increased the evening before the hurricane, and nightlife spending the night after the

---

[301] *Id*.
[302] *See* Daniele Keats Citron, Frank Pasquale, *The Scored Society: Due Process for Automated Prediction*, 89 WASH. L. REV. 1, 17 (2014) (discussing the FCRA).
[303] *See* 15 U.S.C. § 1681(a)(d)(1)(2012).
[304] *See* Citron *supra* note 300.
[305] *Id*.
[306] *See* Mikella Hurley, Julius Adebayo, *Credit Scoring in the Era of Big Data*, 18 YALE J. OF L. & TECH. 148, 152 (2016). *See also* Bobby Allyn, "How the careless errors of credit reporting agencies are ruining people's lives" The Washington Post (September 8, 2016) available at https://www.washingtonpost.com/posteverything/wp/2016/09/08/how-the-careless-errors-of-credit-reporting-agencies-are-ruining-peoples-lives/?utm_term=.a8083de9383b (last accessed June 27, 2017). *See also* "Civil Rights, Big Data, and Our Algorithmic Future, A September 2014 Report on Social Justice and Technology" available at https://bigdata.fairness.io/ (last visited June 27, 2017) (detailing racial bias issues in the government run work-eligibility of potential employees which carries a 20% higher error rate for those who are foreign born as opposed to those born in the United States).
[307] *See* Kate Crawford, *The Hidden Biases in Big Data*, Harvard Business Review (April 1, 2013) available at https://hbr.org/2013/04/the-hidden-biases-in-big-data.
[308] The simplest form of de-identification is the removal of all identifying data such as names, addresses and any other personally identifiable data. *See* Mattioli *supra* note 25 at 567. However, doing so can weaken the effectiveness of the data. *Id*. More sophisticated methods are also used whereby fake values are inserted into the mix of data, secreting the identity or other data of the data population. *Id*. Data analytics assist in mining data including air-sensor feedback, and weather data. *See* Breggin, *supra* note 16 at 10985.
[309] *See* Crawford *supra* note 305.

hurricane.[310] However, more surprisingly the majority of the tweets came from Manhattan, creating an impression that the hurricane significantly impacted Manhattan when in fact it had not.[311] The locations where the hurricane impacted the most had minimal amounts of Twitter messages.[312] Accordingly, if researchers had not thoroughly analyzed the data, the results could have been improperly skewed.[313]

Another interesting example of a potentially erroneous analysis of big data was Boston's attempt to detect and rectify its pothole problems through an application called StreetBump, which passively detects potholes through GPS and acceleration data.[314] The issue with this data collection method was that it did not account for individuals in locations with limited cell phone use or limited cars, occurring in generally lower income areas.[315] Fortunately, the individuals associated with collecting the data recognized this as a possibility and accounted for the discrepancies.[316] These examples demonstrate that without alternate, open sources of verifiable data, erroneous predictions and results are more likely to occur.

4. Self-regulatory Attempts to Rectify Misinformation

Notably, one data broker who recognizes the potential pitfalls of erroneous data collection or interpretation is attempting to combat the issues through a website that allows individuals to log in and correct any errors to their information.[317] It also provides an option for consumers to opt-out of data collection.[318] However, some critics argue that the opt-out mechanism is simply another means by which the data broker giant accesses more information about you instead of truly allowing you to opt-out.[319] In light of the data broker's attempt at transparency and opportunities for consumers to opt-out of the collection of data, there could be some interest in self-regulation by these entities. This might include advanced marketing techniques and trademarks signifying the entities' commitment to transparency and cooperation similar to the "Fair Trade" marketing movement for consumer products.[320]

B. Intentional Manipulation of Big Data

---

[310] *Id.*
[311] *Id.*
[312] *Id.*
[313] *Id.*
[314] *Id.*
[315] *Id.*
[316] *Id.* Crawford also noted Google's failure to accurately predict flu trends and its failure to indicate why it was an outlier. *Id.*
[317] *See* "Optout" available at https://isapps.acxiom.com/optout/optout.aspx (last accessed June 27, 2017).
[318] *Id.*
[319] *See* Will Simonds, "*Acxiom's letting you see the data they have about you (kind of)*" The Online Privacy Blog (September 4, 2013) available at https://www.abine.com/blog/2013/acxioms-letting-you-see-data/ (last visited June 27, 2017). Consumer advocacy group, Stop Data Mining, has compiled a master list of how consumers can opt-out of data collection. *See* "Stop Data Mining.me Opt Out List" available at http://www.stopdatamining.me/opt-out-list/ (last accessed June 27, 2017).
[320] For example, the Fair Trade USA nonprofit organization certifies those products that meet require minimum standards for fair prices, wages, working conditions, environmental and community based protections available at https://fairtradeusa.org/about-fair-trade-usa/who-we-are.

In the context of research, even more troubling than the unintentionally inaccurate data results are those results that have been intentionally manipulated by private sources such as (1) the Facebook emotion experiment, (2) the Facebook voting experiment, and (3) the OkCupid compatibility experiment discussed below. Big data's influence on public opinion has become increasingly concerning. Joe Turow, Cass Sustein and other researchers detailed how commercial entities, like Facebook, can modify their consumers' opinion simply through the types of data Facebook allows its users to see.[321] Potential voters can be recipients of targeted election data based on their fears, interests, or supported causes identified through their online usage.[322] Unlike public researchers who are governed by federal regulations, private researchers are subject to limited privacy restrictions and their own privacy policies.[323] The Federal Policy for the Protection of Human Research Subjects, known as the Common Rule, governs human subject research conducted by public researchers (those that receive federal funding or those who voluntarily comply with its terms for privately funded research).[324] The Common Rule's fundamental policy is that researchers fully and completely inform the subjects about the data researchers gather about them and its use.[325]

A prime example of intentional manipulation of research is Facebook's emotion research where researchers and Facebook exposed negative newsfeeds to certain subscribers and positive newsfeeds to others to determine whether the groups exposed to more negative feeds had more negative postings.[326] Those with the more negative posts demonstrated more negative re-posts to others, and those receiving the positive posts had re-posted more positive material.[327] Both Facebook and the academic researchers faced significant criticism for what the public perceived was unethical human research.[328] The criticism included concerns over the subject pool as it had no age filter and could have included minors without parental informed consent, or nationals from other countries potentially violating international data protection laws,[329] and the researchers did not follow the Common Rule protocols.[330] Federal regulations which provide

[321] *See* Tene *supra* note 69 at 359.
[322] *Id*. at 360 (recognizing that the federal government is using big data to assist in their policy analysis). *See* Breggin *supra* note 16 at 10991. An interesting example includes data that shows a significant draw of power can help law enforcement locate marijuana use. *Id*. at 10992.
[323] *See* James Grimmelmann, *The Law and Ethics of Experiments on Social Media Users*, 13 COLO. TECH. L. J. 219, 226 (2015).
[324] *Id*.
[325] *Id*.
[326] *See* Bambauer *supra* note 286 at 489-91.
[327] *Id*.
[328] *See* Calli Schroeder, *Why Can't We Be Friends? A Proposal for Universal Ethical Standards in Human Subject Research*, 14 COLO. TECH. L. J. 409, 418 (2016). The study sparked controversy regarding ethical standards in research and whether researchers (particularly those working for public universities) should comply with the Common Rule for these types of research projects despite the project being privately funded. *Id*. 700,000 unwitting Facebook subscribers were the subject of the experiment conducted between Facebook and researchers from Cornell University. *Id*.
[329] *Id*. at 412-13.
[330] Federal regulations known as Federal Policy for the Protection of Human Subjects a/k/a the Common Rule require all federally-funded entities or research to follow certain institutional review board standards that examine the risks, minimize those risks, identify the benefits, and compare the reasonableness of the risks to the benefits, and ensure subjects are fully informed and consent, and provide periodic review and monitoring. *Id*. at 412-13.

parameters surrounding human research studies apply to certain categories of government-funded research, and since Facebook was a private entity, such regulations were inapplicable.[331]

In another example of intentional manipulation of data from 2010, researchers hired by Facebook experimented on approximately 61 million Facebook users immediately preceding the 2010 mid-term elections.[332] The experiment divided users into two groups: in one, the researchers showed users a "go vote" message in a plain box; the other group was shown the same box with the addition of thumbnail pictures of their friends who had clicked on "I voted."[333] After the election, the researchers compared the two groups through voting poll records and determined the latter had hundreds of thousands of voters whereas the former group did not.[334] How Facebook conducted this research, and the extent of its users' knowledge of their participation in the experiment is solely within Facebook's control, and exemplifies the need for disclosure obligations.[335]

Moreover, a growing concern among privacy advocates is the advent of psychological targeting through big data.[336] Data brokers know "your age, income, favorite cereal and when you last voted."[337] Companies or politicians can target their marketing efforts to correspond with your psychological profile, e.g. if you are deemed a worrier, data brokers and others may show you ads or news about the dangers of the Islamic State to assist in driving you towards a political candidate or product.[338] This psychological profiling is also known as "emotion analysis" and social media sites conduct this form of profiling.[339] Companies engaged in psychological profiling note that the United States is an easy target as our privacy laws surrounding the data gathered on individuals are minimal unlike the European Union.[340]

For example, OkCupid conducted psychological profiling experiments on subscribers to understand which aspects of their profile were the most relevant.[341] Unbeknownst to its subscribers, OkCupid experimented on 500 of them, telling a group they were incompatible with one another and another group that they were compatible with one another.[342] The experiment noted that when individuals are told they are compatible, they act as if they are, even when they

---

[331] The regulations require the participants to have specific knowledge and consent. *Id*. Although Facebook contends its privacy policy covered the informed consent, it was not until four months after the study that Facebook revised its policy to state: "we may use the data we receive about you . . . for internal operations, including troubleshooting, data analysis, testing, research and service improvement…" *Id*. at 415. Moreover, the study included adolescents and minors, raising red flags concerning mood manipulation and lack of parental knowledge and consent. *Id*. at 412.

[332] *See* Zeynep Tufekci, *Mark Zuckerberg is in Denial* (Nov. 15, 2016) available at www.nytimes.com/2016/11/15/opinion/mark-zuckerberg-is-in-denial.html.

[333] *Id*.

[334] *Id*.

[335] *Id*.

[336] *See* Nicholas Confessore & Danny Hakim, *Data Firm Says "Secret Sauce" Aided Trump: Many Scoff* (Mar. 16, 2017) available at https://www.nytimes.com/2017/03/06/us/politics/cambridge-analytica.html?_r=1.

[337] *Id*.

[338] *Id*.

[339] *Id*.

[340] *Id*.

[341] *See* Grimmelmann *supra* note 321 at 224.

[342] *Id*.

are not.[343] Likewise, those who are told they were incompatible did not seek further contact with the person despite their actual compatibility.[344] Again, OkCupid is a private entity and merely relies on its terms of use to conduct frequent and ongoing research of its subscribers.

Data brokers' experiments reflect a common theme, data brokers and other entities who utilize the public Internet can and do conduct research on their users, whether designed significantly to impact society or otherwise, the effects are the same. Those with access to private data control the future of research and contribute to data inequality.[345]

VII. Impact on Research and the Need for a Cross-Pollination[346] of Data between Data Brokers and the Government

Government agencies protect an individual's privacy through data anonymization and refusal to release discretionary data.[347] For the researcher without access (or resources) to data from brokers, the primary source of research data remains through government agencies' release of data.[348] Data released by government agencies also serves an important societal purpose by providing information to the public.[349] The data released by the government is precisely the type of data that data brokers sell.[350] Without the governmental release of data or the release of data so anonymized as to make the data useless, only those researchers with adequate funding or those with relationships with data brokers will constitute the field of future research. Thus, data inequality (the inability to afford access to unbiased and accurate data) will lead to further income inequality. Without access to unbiased, verifiable data, the public will not know whether what they are seeing, hearing, researching, buying, or voting on, is accurate data or whether it has been intentionally or unintentionally manipulated by data gathered and supplied by the opaque big business of data brokers. If we are told what we think and what is accurate without being able to properly and accurately challenge that data, the public can be further divided between the educated, powerful rich, and the manipulated weakened poor, contributing to negative rent-seeking.

---

[343] *Id*.
[344] *Id*.
[345] While some results and analysis may be unintentionally erroneous, researchers and legal scholars have noted big data research can be used to intentionally discriminate against certain populations. *See* Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security*, 93 TEX. L. REV. 85, 102-03 (2014) (arguing because data can be gathered from a multitude of devices, vendors and others can take action based on this data to your detriment). Once the government releases the data, data brokers often collect the data, analyze it and resell it for a variety of purposes including marketing, credit scoring, or screening job applicants. *See* Borgesius *supra* note 18 at 2092 (noting that household data like whether a smoker lives in the home could be used to decline insurance). Banks and other entities can use big data to assist them in determining the credit worthiness of potential clients. *See* Citron *supra* note 300 at 17 (discussing the FCRA). Companies can segregate data by certain zip codes, disparately impacting lower income individuals.
[346] A cross-pollination of data would necessarily require cooperation between the public and private sector. We contend where information is gathered in one, the other should have the right to access it for certain purposes free of charge.
[347] *See generally,* Altman et al. *supra* note 74.
[348] *Id*.
[349] *Id*.
[350] *Id*.

To combat data inequality and negative rent-seeking, we recommend several potential solutions: (a) legislation requiring data brokers and other online services to provide full and complete disclosure to users regarding the information they collect, its reuse, and potential aggregation and resale, (b) legislation allowing for users to opt-out of data collection and reuse without forgoing use of the data broker's or other online provider's services, (c) legislation requiring data brokers share underlying data with the government and researchers necessary for research in the fields of public welfare, and national security, (d) modification of the government's analysis in discretionary release of information to include evaluating a person's revelation of the same information sought to be released to data brokers or other online providers, (e) encouraging the data broker industry to voluntarily provide disclosure and opt-out mechanisms for users, and (f) encouraging the data broker industry to voluntarily adopt a certification of transparency which could be used as a marketing tool while simultaneously reducing the negative consequences associated with their opaque data process.

We recognize two (2) significant obstacles to the legislative suggestions. First, the current administration has evidenced an intent to eliminate and reduce governmental regulations and has reduced both the FTC's and FCC's authority to regulate the data broker business.[351] Thus, it is highly unlikely any direct legislation or regulation in this area will be forthcoming in the near future. Second, any legislative disclosure or disclaimer obligations must meet the exacting standards of *Citizens United* and *Sorrell*.[352] Regulating data brokers and the data they collect has been met with numerous legal and scholarly challenges ranging from intellectual property rights, contract rights, and constitutional rights.[353] Individuals currently do not have copyright protection to the facts they release to the Internet.[354] Further, there is no fundamental right to privacy in most consumer activity on the Internet, and what privacy does exist, the consumer often relinquishes through consent to the terms of use and service provided by the Internet provider.[355] Some argue that a movement similar to the European Union is relevant and that there is a fundamental right to specific knowledge about what data brokers gather and how others use it, with an affirmative right to opt-out.[356] This issue becomes more worrisome when we

[351] *See generally*, Cecilia Kang, *Congress Moves to Strike Internet Privacy Rules from Obama Era* (Mar. 23, 2017) available at https://www.nytimes.com/2017/03/23/technology/congress-moves-to-strike-internet-privacy-rules-from-obama-era.html?_r=0.

[352] *Citizens United v. Fed. Election Com'n,* 558 U.S. at 370 and *Sorrell v. IMS Health Care, Inc.*, 131 S.Ct. at 2667-68.

[353] *Compare* Fred H. Cate, *Privacy in the Information Age* (1997) (arguing privacy is "an antisocial construct ... [that] conflicts with other important values within the society, such as society's interest in facilitating free expression....") *with* Jane Yakowitz Bambauer, *Is Data Speech?*, 66 STAN. L. REV. 57 (2014) and David Post, *Cyberprivacy, or What I (Still) Don't Get*, 20 TEMP. POL. & CIV. RTS. L. REV. 249, 251 (2011) ("[O]ne person's privacy is very often another person's infringement of the freedom to speak.") and Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 390 (2008) ("Indeed, when it comes to database regulation, many feel that any government regulation of private data flows raises serious First Amendment issues.").

[354] *See e.g.* Feist v. Pub v. Rural 499 US 340, 344 (1991) (finding no right to data analysis of your biomedical data). *See generally* Peter Yu, *The Political Economy of Data Protection*, 84 CHI-KENT L. REV. 777 (2010).

[355] *See* Diana Liebenau, *What Intellectual Property Can Learn from Privacy and Vice Versa*, 30 HARV. J. OF L. TECH 285, 296 (2016) (users often grant a site the nonexclusive, royalty-free, worldwide right to use and license or sublicense their content). Users often do not understand what they are agreeing to and thus their voluntariness is questionable. *See* Daniel Solove, *The FTC & New Common Law of Privacy*, 114 COLUM. L. REV. 583, 667 (2016).

[356] *See* Bradyn Fairclough, *Privacy Piracy: The Shortcomings of the US' Data Privacy Regime & How to Fix It*, 42 J. CORP. L. 461, 479-80 (2016) (arguing FIPs' obligations reflect an acknowledgment that there is a "right" to personal data and discussing the EU's treatment of data protection as a fundamental human right).

consider the ease of purchasing data from data brokers which results in an increased demand for privacy in the release of government data.[357] The idea that data brokers must share their underlying data with the government even for limited topics likely would be met with extreme opposition.[358] Although we believe reducing negative rent-seeking and data inequality are significant governmental interests, any legislation in this area would need to be narrowly tailored and likely would be limited to advising users of the use and aggregation of their data along with their ability to opt-out.

Data brokers' intellectual property concerns with sharing their underlying data could be alleviated through the execution of data use agreements similar to those in federally-funded and restricted research relationships.[359] The proposed data use agreements could require a formal review process comprised of both public and private stakeholders to identify the need for the information and why it is unavailable elsewhere. The agreement could also limit the data's reuse without the data broker's prior consent. If data use agreements or disclosure legislation are not feasible, there are additional mechanisms for furthering data equality, such as tax incentives, public information campaigns on the diminished credibility of research that lacks transparency, and public-private partnerships with data brokers, like Acxiom. Because Acxiom has acknowledged the inaccuracies inherent in data gathering, it may be willing to explore whether more transparency in research is warranted, highlighting the confidence in their data gathering techniques and allowing information derived from them to be challenged and openly corroborated.[360] A trend in favor of those data brokers with transparent data through industry best practices should be promoted-i.e. those that comply could tout their transparency with a certificate and trademark regarding their independence and transparency like "Transparent Data."

Finally, and most importantly, one avenue for reducing data inequality is for the federal government to assess privacy concerns in a broader context, limiting the discretionary FOIA denials.[361] For these and other data releases, the government should consider whether the individual whose privacy is at issue has otherwise released the information through other commercial online means. In this regard, researchers may have more avenues to access accurate and transparent data.

---

[357] *See* Dwork *supra* note 229.

[358] It is likely that businesses will challenge a forced disclosure requirement as they have done with individual state legislation regarding genetically modified labels and release of toxic chemicals. *See* Gary D. Bass, *Big Data Government Accountability: An Agenda for the Future*, 11 I/S: J. L. & POL'Y FOR INFO. SOC'Y 13, 21-23 (2015) (arguing for a proactive disclosure requirement for the government to follow). *Id.* at 37. *See also*, Bradford W. Hesse, Richard P. Moser, William T. Riley, *From Big Data to Knowledge in the Social Sciences*, 659 ANNALS AM. ACAD. POL. & SOC. SCI. 16, 19-21 (2016) (discussing federally-funded research often requires full publication of the research data within 12 months of publication).

[359] *See e.g.* Practice Guide: "Data Use Agreement, Department of Health and Human Services" available at https://www.hhs.gov/ocio/eplc/EPLC%20Archive%20Documents/55-Data%20Use%20Agreement%20(DUA)/eplc_dua_practices_guide.pdf (last accessed June 27, 2017).

[360] *See* Simonds, *supra* note 317.

[361] *But see*, David E. Pozen, *Deep Secrecy*, 62 STAN. L. REV. 257 (2010) (discussing the difficulty in accessing governmental information particularly where the existence of the information is hidden and examining the theories behind secrecy). *See also*, Samaha, *supra* note 33 (detailing the negative consequences surrounding one's ability to access too much public information).

VIII. Conclusion

It is unreasonable to require privacy in government data maintenance while private data sources are free to enjoy immense benefits without commensurate privacy obligations. Inevitably, this form of imbalanced burden shifting will result in *data inequality* whereby those with the resources have access to data, and the rest of the public will have little or no access to meaningful data. We argue for a reduction in data inequality and the proper balancing of the government's privacy obligations compared to the data brokers' infinite access to data. Although regulatory reform in this area is unlikely during the Trump administration, the government can ensure it releases more records within its discretionary release capabilities, can incentivize data brokers to release necessary data for research, and could implement an informational campaign regarding the information the data brokers gather and the lack of credibility any research has without the ability to cross-check the underlying data provided by data brokers. Ultimately, a combination of solutions similar to those recommended herein would reduce the growing data inequality and limit any concomitant, negative rent-seeking effects.