Maine Law Review

Volume 66
Number 2 *Symposium: Who's Governing Privacy? Regulation and Protection in a Digital Era*

Article 3

June 2014

The Glass House Effect: Big Data, The New Oil, and the Power of Analogy

Dennis D. Hirsch

THE GLASS HOUSE EFFECT: BIG DATA, THE NEW OIL, AND THE POWER OF ANALOGY

Dennis D. Hirsch

THE GLASS HOUSE EFFECT: BIG DATA, THE NEW OIL, AND THE POWER OF ANALOGY

Dennis D. Hirsch[*]

I. INTRODUCTION

One hears with some frequency today that "data is the new oil."[1] Recently, Virginia Rometty, IBM's Chief Executive Officer, updated the phrase, explaining that *Big* Data is the new oil.[2] Most people who use the analogy do so in order to convey Big Data's tremendous value. Data is an essential resource that powers the information economy much like oil has fueled the industrial economy. Big Data[3] promises a plethora of new uses—the identification and prevention of pandemics,[4] the emergence of new businesses and business sectors,[5] the improvement of health care quality and efficiency,[6] and enhanced protection of the environment,[7] to name but a few—just as oil has generated useful plastics, petro-chemicals, lubricants, and

---

* Geraldine W. Howell Professor of Law, Capital University Law School. The Author would like to thank Professor Paul Ohm for suggesting the idea for this paper, and for early discussions that helped to shape it. Unless otherwise indicated, the author alone is responsible for the paper's content. Portions of this Article were included in a paper presented at the Future of Privacy Forum and the Stanford Center for Internet & Society's workshop, *Big Data and Privacy: Making Ends Meet*, *available at* http://www.futureofprivacy.org/big-data-privacy-workshop-paper-collection.

1. *See, e.g.*, Clive Humby, Address at the ANA Senior Marketer's Summit at the Kellogg School (2006); Michael Palmer, *Data is the New Oil*, ANA MKTG. MAESTROS (Nov. 3, 2006, 5:43 AM), http://ana.blogs.com/maestros/2006/11/data_is_the_new.html; Personal Data: The "New Oil" of the 21st Century, Panel Discussion of the World Economic Forum on Europe & Central Asia (June 9, 2011), *available at* http://www.weforum.org/sessions/summary/personal-data-new-oil-21st-century.

2. Maria Deutscher, *IBM's CEO Says Big Data is Like Oil, Enterprises Need Help Extracting the Value*, SILICON ANGLE (Mar. 11, 2013), http://siliconangle.com/blog/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value. *See also* VICTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK 16 (2013); Julie Brill, Comm'r, Fed. Trade Comm'n., Sloan Cyber Security Lecture: A Call to Arms: The Role of Technologists in Protecting Privacy in the Age of Big Data 1 (Oct. 23, 2013), *available at* http://www.ftc.gov/sites/default/files/documents/public_statements/call-arms-role-technologists-protecting-privacy -age-big-data/131023nyupolysloanlecture.pdf (stating that Big Data is the "new oil") [hereinafter Brill Lecture]; Sam Pfeifle, *Big Data=Big Oil?,* IAPP: THE PRIVACY ADVISOR (Apr. 1, 2013), https://www.privacyassociation.org/publications/2013_03_20_big_data_big_oil.

3. "The term 'Big Data' refers to datasets whose size is beyond the ability of typical database software tools to capture, store, manage and analyze." ANN CAVOUKIAN & JEFF JONAS, INFO. & PRIVACY COMM'R (ONTARIO, CAN.), PRIVACY BY DESIGN IN THE AGE OF BIG DATA 3 (June 8, 2012), *available at* http://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf.

4. MAYER-SCHÖNBERGER & CUKIER, *supra* note 2, at 2-3.

5. *Id.* at 3.

6. JAMES MANYIKA ET AL., MCKINSEY GLOBAL INST., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION AND PRODUCTIVITY 2 (May 2011), *available at* http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_exec_summary.ashx.

7. Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 248 (2013).

gasoline. Big Data "is becoming a significant corporate asset, a vital economic input, and the foundation of new business models. It is the oil of the information economy."[8]

This Article looks at the analogy in a different way, one not yet developed in the scholarly literature. It examines the underside of the 'Big Data is the new oil' comparison. Oil certainly has many productive uses, but it also leads to oil pollution. Big Data is similar. It produces tremendous benefits, but simultaneously generates significant privacy injuries.[9] As the data sets get larger, the threat grows as well.[10] Big Data is like a massive oil tanker navigating the shoals of hackers, criminals and human error. It can make us smarter and wealthier and our lives better. However, like oil, it can also harm us. Environmental law has developed ways to reduce oil pollution. This Article draws on this environmental law success story to identify ways that law and policy can protect privacy in the era of Big Data.

The Article begins by developing the analogy between oil pollution and privacy injuries. Oil pollutes in two principal ways. It spills, and so despoils beaches, coastlines and waters. It also produces carbon emissions and so contributes to the greenhouse effect and climate change. Big Data creates analogous privacy injuries. Like oil, it spills. Data security breaches—such as the 2013 Target Thanksgiving and Christmas shopping season breach, in which hackers gained access to an estimated 70-110 million customer names, credit and debit card numbers, expiration dates and security codes[11]—cause broad harm, much as oil spills create wide-spread damage. Big Data's privacy impacts are also analogous to carbon emissions and climate change. Oil combustion contributes to a growing layer of greenhouse gases that traps the sun's heat, causes climate change, and so makes the physical environment less hospitable for humans and other forms of life.[12] In a similar way, the producers of Big Data are generating layer upon layer of personal information. This build-up increases the hot glare of public scrutiny and so makes the social environment less conducive to the growth of the human personality which requires a degree of shade and shelter in which to

8. MAYER-SCHÖNBERGER & CUKIER, *supra* note 2, at 16.

9. STANLEY R.M. OLIVEIRA & OSMAR R. ZAÏANE, TOWARDS STANDARDIZATION IN PRIVACY-PRESERVING DATA MINING § 3.1 (2004), *available at* http://www.agencia.cnptia.embrapa.br/Repositorio/dm-ssp04ID-9ZWpbbVpUP.pdf (explaining that, while "data mining can be extremely valuable in many applications (e.g., business, medical analysis, etc[.]), it can also, in the absence of adequate safeguards, violate informational privacy").

10. This is true, first, because more personal data is released. It is further true because it is easier to re-identify large de-identified data sets, than to re-identify small ones.

11. *See Chronology of Data Breaches,* PRIVACY RTS. CLEARINGHOUSE, http://www.privacyrights.org/sites/privacyrights.org/files/static/Chronology-of-Data-Breaches_-_Privacy-Rights-Clearinghouse.pdf (last visited Feb. 8, 2014) (explaining that Target "customers who used a payment card at any of Target's stores nationwide between November 27, 2013 and December 15, 2013 may have had their payment card information copied for fraudulent purposes. Customer names, credit or debit card numbers, card expiration dates, and card security codes were taken and have appeared on the black market.")

12. CHRIS WOLD, DAVID HUNTER, & MELISSA POWERS, CLIMATE CHANGE AND THE LAW 5 (2009) (explaining that the burning of fossil fuels releases carbon dioxide which is "by far the most important" man-made greenhouse gas).

flourish. This is not the greenhouse effect, but the *glass house effect*, since it gives each of us the sense that we are living in a glass house. Climate change is a good analogy for a transformation so profound that it is at once happening all around us and, at the same time, difficult to grasp and identify.

In their Harvard Business Review Article, *How Strategists Really Think: Tapping the Power of Analogy*, Professors Gavetti and Rivkin explain that analogies can not only help us understand contemporary problems; they can also enable us to develop solutions to them.[13] Strategic thinkers begin by comparing a new problem to a prior one for which they have a remedy.[14] Then they take the solution to the prior problem, adapt it to the new context, and so develop ideas about how to address the current issue.[15] The "Big Data is the new oil" analogy is powerful in this way. Society has developed legal and policy solutions to oil pollution. We should be able to take these measures—the solutions to the familiar problem—translate them into the privacy realm, and so gain insight into how to reduce Big Data's impacts on privacy.

That is what the second portion of this Article attempts to do. It explains how the Clean Water Act[16] and the Oil Pollution Act[17] succeeded in reducing oil tanker spills. It takes these strategies, adapts them for the privacy realm, and so produces a set of legal and policy recommendations for decreasing data spills. It then turns to climate change policy—particularly laws and policies designed to promote clean energy innovation—and translate it into a federal government strategy for promoting technologies that can allow us to achieve the many benefits of Big Data, while reducing its harmful effects on privacy. This will mitigate the glass house effect.

This project is important, not only for the protection of privacy, but for the future of Big Data and data analytics itself. Consider the following example: New York, Oakland, and other cities have been collecting massive amounts of surveillance camera data and mining it for law enforcement purposes.[18] This promises to reduce crime and increase personal safety.[19] It also constitutes a major business opportunity for IBM and Microsoft, the providers of this Big Data

---

13. *See* Giovanni Gavetti & Jan W. Rivkin *How Strategists Really Think: Tapping the Power of Analogy*, HARV. BUS. REV., Apr. 1, 2005, at 1.

14. *Id.* at 2.

15. *Id.* As Gavetti and Rivkin explain it, strategic, analogical reasoning involves

> a novel problem that has to be solved or a new opportunity that begs to be tapped; a specific prior setting that managers deem to be similar in its essentials; and a solution that managers can transfer from its original setting to the unfamiliar context. When managers face a problem, sense "Ah, I've seen this one before," and reach back to an earlier experience for a solution, they are using analogy.

Id.

16. 33 U.S.C. §§ 1251-1387 (2012). Congress added the Clean Water Act's oil spill provisions in the Water Quality Improvement Act of 1970, Pub. L. No. 91-224, 84 Stat. 91 (1970), and in the Federal Water Pollution Control Act Amendments of 1972, Pub. L. No. 92-500, 86 Stat. 816 (1972).

17. Oil Pollution Act of 1990, Pub. L. No. 101-380, 104 Stat. 484 (1990).

18. *See* Somini Sengupta, *Privacy Fears Grow as Cities Increase Surveillance*, N.Y. TIMES, Oct. 13, 2013, at A1, *available at* http://www.nytimes.com/2013/10/14/technology/privacy-fears-as-surveillance-grows-in-cities.html.

19. *Id.*

service.[20] However, the program has come under fire for compiling "data about the everyday movements and habits of law-abiding residents, raising legal and ethical questions about tracking people so closely."[21] At least one city has placed a moratorium on the use of some surveillance devices such as license plate readers.[22] As this example shows, Big Data promises many beneficial uses and applications. It offers crime reduction, improved health and safety, new services and industries, greater efficiency, and much more. Yet, left unaddressed, its alarming impacts on individual privacy could provoke a political backlash that will inhibit and limit its use.[23] Some organizations, anticipating such a reaction, have already "clamped down on their sensitive data, uncertain about what, if anything, they can release without jeopardizing the privacy of individuals."[24] "If privacy concerns are not adequately addressed, they may stall or disrupt the deployment of new technologies that offer many potential economic and quality-of-life benefits to consumers."[25] In order to unlock the great potential of Big Data, society must find ways to address and prevent the privacy threats that it poses.[26] "Big data is the new oil" is a powerful analogy that can suggest strategies for achieving this.

II. BIG DATA IS THE NEW OIL (POLLUTION)

Big Data and data analytics will create many important benefits for society. The positive side of the "Big Data is the new oil" analogy is, in many ways, an accurate comparison. However, it is vital also to appreciate the negative dimension of the analogy—the comparison between Big Data's privacy impacts and oil pollution. As explained above, it is only by doing so that society will be able to unlock Big Data's great potential. The positive and negative dimensions of the analogy are linked: in order to have the first, one must also explore and address the second.

The oil-based economy generates two main types of pollution: oil spills that despoil waters, beaches and coastlines; and carbon emissions that contribute to climate change.[27] Each of these forms of oil-based pollution is analogous to Big

---

20. *See id.*

21. *Id.*

22. *Id.*

23. *See* CAVOUKIAN & JONAS, *supra* note 3, at 3 (explaining that "technological advances improve our ability to exploit Big Data, potential privacy concerns could stir a regulatory backlash that would dampen the data economy and stifle innovation"); Omer Tene & Jules Polonetsky, *Privacy in the Age of Big Data: A Time for Big Decisions*, 64 STAN. L. REV. ONLINE 63, 63 (2012) (suggesting the same).

24. Erica Klarreich & Simons Science News, *Privacy by the Numbers: A New Approach to Safeguarding Data*, SCI. AM. (Dec. 31, 2012), http://www.scientificamerican.com/article/privacy-by-the-numbers-a-new-approach-to-safeguarding-data.

25. DANIEL CASTRO, INFO. TECH. & INNOVATION FOUND., THE NEED FOR AN R&D ROADMAP FOR PRIVACY 1 (Aug. 2012), *available at* http://www2.itif.org/2012-privacy-roadmap.pdf.

26. *See* WORLD ECON. FORUM, UNLOCKING THE VALUE OF PERSONAL DATA: FROM COLLECTION TO USAGE 11-13 (Feb. 2013), *available at* http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf (explaining that better privacy governance will allow the continued and expanded use of Big Data).

27. The use of oil and oil-derivatives such as gasoline contribute to other types of pollution such as the release of volatile organic compounds that produce ground-level ozone (i.e., smog), leaks from

Data's privacy impacts.

A. Data Spills are Like Oil Spills

The extraction, transportation and storage of oil and oil-based derivatives (such as gasoline or home heating oil) inevitably lead to oil spills. The 1989 Exxon Valdez spills that released 11 million gallons of crude oil into Alaska's Prince William Sound,[28] and the 2010 Deepwater Horizon oil gusher in the Gulf of Mexico,[29] are dramatic examples, but hardly the only ones. There have been many, many other oil spills, both on land and at sea. Oil spills can injure coastlines, beaches, fish, marine ecosystems, and water quality.[30] They can also damage commercial industries, such as fishing or tourism. In all of these ways, oil spills cause great damage.[31]

Data spills do too. Since 2005, there have been more than four thousand reported data breach incidents,[32] an estimate that likely undercounts the true number.[33] Large-scale, recent incidents include the 2013 Target Thanksgiving and Christmas shopping season security breach that released an estimated 70-110 million records;[34] the 2007 TJX Company's international release of an estimated 100 million records; [35] and the 2011 Sony release of 101 million records, including over 12 million unencrypted credit card numbers.[36]

Data brokers[37] and other users of Big Data[38] contribute substantially to the

---

underground storage tanks, and other forms of air and land pollution. Oil spills and carbon emissions are two of the most significant forms of oil-based pollution and will be the focus here.

28. *Exxon Valdez Spill Profile*, U.S. EPA, http://www.epa.gov/oem/content/learning/exxon.htm (last updated Jan. 16, 2014).

29. *Deepwater Horizon Oil Spill,* WIKIPEDIA, http://en.wikipedia.org/wiki/Deepwater_Horizon_oil_spill (last visited Jan. 29, 2014).

30. FRED BOSSELMAN, ET AL., ENERGY, ECONOMICS AND THE ENVIRONMENT: CASES AND MATERIALS 320, 333 (3d ed. 2010).

31. E. A. Barry-Pheby, *The Growth of Environmental Justice and Environmental Protection in International Law: In the Context of Regulation of the Arctic's Offshore Oil Industry,* 13 SUSTAINABLE DEV. L. & POL'Y 48 (2013).

32. *See Chronology of Data Breaches, supra* note 11. In many spills, the number of records released is unknown. The estimate does not include these incidents. *See id.* It further excludes spills, of which there may be many, that do not release Social Security Numbers or financial information. *See id.*

33. This estimate includes only those breaches that involved social security numbers or financial information, and that *excludes* the many breaches for which the number of records released is not known. *See id.*

34. *See id*. (explaining that Target "customers who used a payment card at any of Target's stores nationwide between November 27, 2013 and December 15, 2013 may have had their payment card information copied for fraudulent purposes. Customer names, credit or debit card numbers, card expiration dates, and card security codes were taken and have appeared on the black market.")

35. *See id*. (describing the 2003-2006 security breach at TJX that released data about "credit card, debit card, check, and merchandise return transactions.")

36. *See id*. (describing the cybercriminal attack on the Sony data center in San Diego).

37. *See* Jessica Rich, Dir., Bureau of Consumer Protection, Fed. Trade Comm'n, Privacy Today and the FTC's 2014 Privacy Agency, Address Before the Int'l Ass'n of Privacy Prof'ls (Dec. 6, 2013), *available at* http://www.ftc.gov/sites/default/files/documents/public_statements/privacy-today-ftcs-

problem.[39] For example, Acxiom, a major data broker, possesses an average of 1,500 data points on 500 million active consumers worldwide, including the majority of American adults.[40] A 2003 Acxiom data breach released an estimated 1.6 billion records containing personal information.[41]

If data is the new oil, then these data releases are the new oil spills. In fact, they have come to be known as "data spills."[42] The analogy runs deep. Just like oil spills, data spills cause different types of damages. These include identity theft, in which criminals use released personal information to impersonate the individual and withdraw money, open credit cards, take out loans, or make purchases in that person's name;[43] the increased *risk* of identity theft, which causes the victims to incur prevention costs (e.g., paying for credit monitoring) and to experience worry

---

2014-privacy-agency/131206privacytodayjrich.pdf (identifying data brokers as a primary example of an industry that uses Big Data).

38. *See generally* STAFF REPORT FOR CHAIRMAN ROCKEFELLER, OFFICE OF OVERSIGHT & INVESTIGATIONS: MAJORITY STAFF, S. COMM. ON COMMERCE, SCI., & TRANS., A REVIEW OF THE DATA BROKER INDUSTRY: COLLECTION, USE AND SALE OF CONSUMER DATA FOR MARKETING PURPOSES (Dec. 18, 2013), *available at* http://www.commerce.senate.gov/public/?a=Files.Serve&File_id=0d2b3642-6221-4888-a631-08f2f255b577 (explaining that contemporary data brokers use extremely large amounts of personal data).

39. The FTC defines data brokers as "companies that collect information, including personal information about consumers, from a wide variety of sources for the purpose of reselling such information to their customers for various purposes, including verifying an individual's identity, differentiating records, marketing products, and preventing financial fraud." FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE 68 (2012), *available at* http://www.ftc.gov/reports/protecting-consumer-privacy-era-rapid-change-recommendations-businesses-policymakers.

40. *See* Natasha Singer, *You For Sale,* N.Y. TIMES, June 17, 2012, at BU1, *available at* http://www.nytimes.com/2012/06/17/technology/acxiom-the-quiet-giant-of-consumer-database-marketing.html?pagewanted=all&_r=1&.

41. *Acxiom Data Breaches*, PSEUDO-FLAW.NET, http://pseudo-flaw.net/content/acxiom-data-breaches (last visited Jan. 29, 2014). Two data breaches at ChoicePoint, in 2003 and 2005, released the personal financial records of 163,000 consumers, *ChoicePoint Settles Data Security Breach Charges; To Pay $10 Million in Civil Penalties, $5 Million for Consumer Redress,* FED. TRADE COMM'N (Jan. 26, 2006), http://www.ftc.gov/news-events/press-releases/2006/01/choicepoint-settles-data-security-breach-charges-pay-10-million, and resulted in at least 800 documented cases of identity theft, *see* Rich, *supra* note 37. LexisNexis, Dun & Bradstreet, Kroll Background America Inc., and the National White Collar Crime Center, who rank amongst the largest consumer and business data aggregators, recently revealed that they too experienced a large data spill. *Data Broker Giants Hacked by ID Theft Service,* KREBSONSECURITY (Sept. 13, 2013), http://krebsonsecurity.com/2013/09/data-broker-giants-hacked-by-id-theft-service; *Data Broker Hackers Also Compromised NW3C*, KREBS ON SEC. (Oct. 13, 2013) http://krebsonsecurity.com/2013/10/data-broker-hackers-also-compromised-nw3c.

42. *See, e.g.*, Brett Glass, *Tower Records Suffers Massive Data Spill*, EXTREMETECH (Dec. 12, 2002, 5:51 AM), http://www.extremetech.com/extreme/52810-tower-records-suffers-massive-data-spill.

43. Victims of identity theft must often spend considerable amounts of money and time contesting the charges or liabilities, restoring their credit, and reclaiming their identity, if they can fully achieve this at all. Some analysts distinguish between identity fraud and identity theft. TERRI CULLEN, THE WALL ST. JOURNAL. COMPLETE IDENTITY THEFT GUIDEBOOK: HOW TO PROTECT YOURSELF FROM THE MOST PERVASIVE CRIME IN AMERICA 11-12 (2007). Identity fraud is when a criminal uses stolen sensitive personal information to steal money from bank accounts, open credit lines using the victim's information, or steal from existing credit card accounts. *Id* at 20. Identity theft is when a criminal uses the stolen, sensitive, personal information to actually impersonate the victim. *Id.*

and stress; and the release of sensitive data (e.g., sexual orientation, HIV status), which can cause the victims to experience acute embarrassment and/or stigma. Oil and data spills are also similar in that each tends to impose small harms on a large number of people. This creates both collective action and free-rider problems for any tort claimants seeking damages for such injuries. The availability of class actions notwithstanding, these common features of oil and data spills frequently frustrate tort remedies and may justify some type of statutory response. Finally, in both oil and data spills, size matters. Just as large oil spills generally cause more damage than small ones, so *big* data spills frequently create more harm than smaller releases. They affect more people. They also increase the chance that hackers and criminals will be able to re-identify purported "anonymized" databases since larger amounts of data increase the probability of correlation with other, identified data, and so facilitate re-identification.[44] For all of these reasons, data spills—particularly *big* data spills—are analogous to oil spills.

B. The Glass House Effect

Big Data's privacy injuries are also analogous to the second principal type of oil-based pollution: carbon emissions and climate change. The combustion of oil and other fossil fuels has generated a layer of carbon dioxide in the upper atmosphere that allows in the sun's rays but then traps its heat.[45] This "greenhouse effect" contributes to global warming and climate change, disturbs ecosystems and weather patterns, and so makes the earth less hospitable for human and natural life.[46] In much the same way, the information economy, and Big Data technologies in particular, are generating a mass of data that is expanding at an exponential rate. According to one estimate, ninety percent of the world's data, from the beginning of human history until the present, has been produced in just the past two years.[47] This total is projected to double every two years for the foreseeable future.[48] Data is growing even faster than atmospheric carbon.

Just as the accumulation of greenhouse gases traps the sun's heat, so the accumulation of data concentrates the hot glare of public scrutiny. It reveals and shines a light on personal information about medical conditions, sexual orientation, political interests, intellectual explorations, and all other manner of sensitive, personal information. This can make us more likely to conform, and less likely to

---

44. *See* CAVOUKIAN & JONAS, *supra* note 3, at 3 (explaining that "Big Data can increase the risk of re-identification").

45. ROBERT PERCIVAL ET AL., ENVIRONMENTAL REGULATION: LAW, SCIENCE, AND POLICY 1207-09 (7th ed. 2013); *see also* INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE, CLIMATE CHANGE 2007: SYNTHESIS REPORT (2008), *available at* https://www.ipcc.ch/pdf/assessment-report/ar4/syr/ar4_syr.pdf [hereinafter IPCC].

46. IPCC, *supra* note 45, at 26; PERCIVAL ET AL., *supra* note 45, at 1207-09.

47. *Big Data, for Better or Worse: 90% of World's Data Generated Over Last Two Years*, SCI. DAILY (May 22, 2013), http://www.sciencedaily.com/releases/2013/05/130522085217.htm; MAYER-SCHONBERGER & CUKIER, *supra* note 2, at 151 ("The size and scale of data collection will increase by leaps and bounds as storage costs continue to plummet and analytic tools become ever more powerful.")

48. Steve Lohr, *The Age of Big Data*, N.Y. TIMES, Feb. 12, 2012, at SR1, *available at* http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html.

experiment, explore, and so to find and become who we want or need to be. In so doing, Big Data makes the social world less hospitable for the full flowering and development of the human personality which requires a degree of shade and shelter in order to flourish.[49] As Julie Cohen has so insightfully put it, it creates a "subtle yet fundamental shift in the content of our character, a blunting and blurring of rough edges and sharp lines. . . . The condition of no-privacy threatens not only to chill the expression of eccentric individuality, but also, gradually, to dampen the force of our aspiration to it."[50]

This is not the greenhouse effect, but the *glass house effect*. Big Data is creating a world in which each of us is increasingly living in a glass house where many of our most personal features—our relationships, anxieties, political beliefs, interests, sexual desires, location, purchases, criminal and financial history, and much else—are observable by others. One need only think about the National Security Agency's collection and storage of who we call and what we do on the Web; or the fact that Google accesses and reads our Gmail messages; or that mobile phone apps invisibly collect our location and contact information, even when they are not at all relevant to the service that the app provides, to get a sense of what is happening. Just as the greenhouse effect alters the earth's climate in ways that make it less friendly to natural ecosystems, so the glass house effect changes the *social* climate in ways that make it far less conducive to personal development. If this trend continues, we will pass on to our children a depleted ecosystem for the cultivation of the human personality. It is vital to human development, creativity and innovation that we prevent this.

It is equally vital to future of Big Data. As mentioned above, Big Data and data analytics possess a truly great potential to enhance human health, quality of life, the environment, economic prosperity and much more.[51] Yet this activity's harmful by-products—the privacy injuries that it creates—are already throwing up obstacles for the field. If left unaddressed, they may well produce a social and political backlash against Big Data, one that could significantly reduce the use of data analytics and prevent it from fully making its valuable contributions. In order to unlock Big Data's potential and achieve its great benefits, it is essential simultaneously to address and mitigate the privacy harms that it creates.[52] Those who care about privacy, and those who wish to advance Big Data and data analytics, should all be interested in finding solutions to Big Data's privacy impacts.

---

49. Julie E. Cohen, *Examined Lives: Informational Privacy and the Subject As Object*, 52 STAN. L. REV. 1373, 1424 (2000) (explaining that "autonomy in a contingent world requires a zone of relative insulation from outside scrutiny and interference—a field of operation within which to engage in the conscious construction of self.")

50. *Id.* at 1426.

51. *See* CAVOUKIAN & JONAS, *supra* note 3; *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 4; *See* MANYIKA ET AL, *supra* note 6; *see* Tene & Polonetsky*, supra* note 7 and accompanying text.

52. Brill Lecture, *supra* note 2 (discussing the operation of Big Data in a way that "respects consumer privacy and engenders consumer trust, allowing big data to reach its full potential to benefit us all.")

III. Oil Pollution Law and Policy as a Model

This takes us back to Gavetti and Rivkin and the power of analogy.[53] The preceding section has argued that Big Data's privacy injuries (the new problem) are analogous to oil pollution (the familiar one). The next step would be take the approaches that society has employed to address oil pollution, translate them into the realm of Big Data and privacy, and so develop strategies for reducing Big Data's privacy impacts.

As described above, the oil and privacy problems are analogous in two distinct ways: oil spills are analogous to data spills; and the increase in greenhouse gases is similar to the accumulation of Big Data. When it comes to oil, these two sorts of pollution are distinct and addressing them requires different strategies. It therefore makes sense to separate out the two types of privacy harms associated with Big Data—data spills, and the glass house effect—and explore separately what environmental regulation can teach us about addressing each of them.

A. The History of Oil Pollution Law Suggests Ways to Reduce Data Spills

At the beginning of the oil era, in the 19th Century, the law supported those who produced and transported this valuable substance and insulated them, to some degree, from large-scale, potentially uninsurable liability for the damage that their activities caused.[54]

To begin with, tort law required victims of an oil spill to demonstrate that the spiller had acted negligently—a difficult task in a risky field where even those who took due care could experience accidents and spills.[55] Further, the law limited the types of damages that were legally cognizable. Maritime tort law recognized property damage from oil spills, but not injuries to fishing, tourism and other non-property-based commercial injuries.[56] Finally, as if tort liability were not yet sufficiently constrained, Congress passed the Limitation on Liability Act of 1851, which capped oil spill damages at the value of the vessel and freight remaining after the accident.[57] Congress intended this statute to facilitate the transportation of an otherwise uninsurable, yet economically critical, cargo. Over time, it came to produce patently absurd results. For example, under the terms of the Act the 1967 wreck of the Torrey Canyon oil tanker, which spilled over 100,000 tons of crude oil into the English channel and despoiled 100 miles of French and British coasts, would have resulted in only $50 in damages—the value of the sole remaining

---

53. *See* Gavetti & Rivkin, *supra* note 13 and accompanying text.

54. *See generally* Mark A. White, *The 1851 Shipowners' Limitation of Liability Act: Should the Courts Deliver the Final Blow?*, 24 N. Ill. U. L. Rev. 821 (2004) (explaining how legal doctrines supported the growth of the nascent American oil transportation industry).

55. *Id.*

56. *See, e.g.*, Robins Dry Dock & Repair Co. v. Flint, 275 U.S. 303, 307 (1927); Louisiana v. M/V Testbank, 752 F.2d 1019, 1023 (5th Cir. 1985).

57. Limitation of Liability Act of 1851, 46 U.S.C. app. § 183 (1984) (current version at 46 U.S.C. §§ 30501-30512 (2012)).

lifeboat.[58]

Congress began to correct the situation in the 1970 and 1972 Amendments to the Clean Water Act.[59] Following the massive Exxon Valdez oil spill, it took even more vigorous action in the 1990 Oil Pollution Act.[60] Together, these statutes rewrite oil pollution law. They allow the government to clean up an oil spill and bring an action against the responsible party to recoup the clean-up costs,[61] thereby reducing the collective action and free-rider problems that undermine private tort actions. They recognize new causes of action for damage to economic, as opposed to simply property-based, interests.[62] For example, it allows commercial fishermen and owners of businesses that rely on beach tourism to sue for the damage that an oil spill caused to their enterprise.[63] Although the statutes do not expressly address the point, courts have interpreted them as rejecting the negligence-based tort regime and creating strict liability for defendants with respect to oil removal and clean-up costs.[64] The statutes greatly increased the amount of damages for which oil spill defendants could be held liable.[65] Today, vessels over 3,000 tons can face liability of up to $22,000,000 per incident.[66] Finally, the Oil Pollution Act requires all new oil transportation vessels operating in U.S. waters to employ a double hull design that greatly decreases the chance of an oil spill.[67] Taken together, these

---

58. Jeffrey D. Morgan, *The Oil Pollution Act of 1990: A Look at its Impact on the Oil Industry,* 6 FORDHAM ENVTL. L. REV. 1, 2 (1994).

59. Water Quality Improvement Act of 1970, Pub. L. No. 91-224, 84 Stat. 91 (1970) (codified as amended in scattered sections of 33 U.S.C.); Federal Water Pollution Control Act Amendments of 1972, Pub. L. No. 92-500, 86 Stat. 816 (1972).

60. Oil Pollution Act of 1990, Pub. L. No. 101-380, 104 Stat. 484 (1990).

61. Water Quality Improvement Act of 1970, §§ 11(c)(1), (f), 84 Stat. at 93 (codified as amended at 33 U.S.C. §§ 1321 (c)(1), (f) (2012).; *see also* Kenneth M. Murchison, *Liability Under the Oil Pollution Act: Current Law and Needed Revisions*, 71 LA. L. REV. 917, 920 (2011).

62. Oil Pollution Act of 1990, Pub. L. 101-380, § 1002(b)(2)(E), 104 Stat. 484, 490 (1990) (codified at 33 U.S.C. § 2702(b)(2)(E) (2012)) (allowing claims for "[d]amages equal to the loss of profits or impairment of earning capacity due to the injury, destruction, or loss of real property, personal property or natural resources, which shall be recoverable by any claimant."); *see also* Murchison, *supra* note 61.

63. *See* John C. P. Goldberg, *Liability for Economic Loss in Connection with the Deepwater Horizon Spill*, 30 MISS. C. L. REV. 335 (2011) (describing newly available claims); Antonio J. Rodriguez and Paul A.C. Jaffe, *The Oil Pollution Act of 1990*, 15 TUL. MAR. L.J. 1 (1990) (same); In re Settoon Towing LLC, 2009 WL 4730969 (E.D. La. 2009) (allowing plaintiffs' claim for loss of profits to go forward); Dunham-Price Group, LLC v. Citgo Petroleum Corp., 2010 WL 1285446 (W.D. La. 2010) (allowing a pure economic claim for loss of use, increased expense, business interruption and related damages under the OPA); FGDI LLC v. M/V Lorelay, 193 Fed. App'x. 853 (11th Cir. 2006) (allowing a grain elevator owner to bring his claim based on the delay that an oil spill caused his business).

64. Murchison, *supra* note 61, at 922. *See, e.g.*, In re Oriental Republic of Uru., 821 F. Supp. 928, 931 (D. Del. 1992); Total Petroleum, Inc. v. United States, 12 Cl. Ct. 178, 180 (1987); United States v. M/V Big Sam, 681 F.2d 432, 440 (5th Cir.1982), *cert. denied*, 462 U.S. 1132 (1983).

65. Oil Pollution Act of 1990, § 1004(a)(1), 104 Stat. at 491-92 (codified as amended at 33 U.S.C. § 2704(a)(1) (2012)). The Oil Pollution Act of 1990 raised the limits to the greater of $1,200 per ton, or $10,000,000 for a vessel greater than 3,000 tons, or $2,000,000 for a smaller vessel. *Id.*

66. Coast Guard and Maritime Transportation Act of 2006, Pub. L. 109-241, § 603(a)(1), 120 Stat. 516, 553 (2006) (codified at 33 U.S.C. § 2704(a)(1) (2012)) (raising the liability amounts).

67. Oil Pollution Act of 1990, § 4115(a), 104 Stat. at 517-18 (codified as amended at 46 U.S.C. § 3703a (2012)).

strategies have significantly improved the oil spill problem. "The Oil Pollution Act of 1990 is widely viewed as an enormous success. It is credited with improving the safety of oil tankers operating in U.S. waters and its double hull requirement has now been adopted internationally."[68]

This environmental law success story holds important lessons for Big Data. As with the early laws governing the oil industry, today's doctrines appear to favor the production, storage and transfer of the "new oil." Plaintiffs alleging damages from data security breaches must generally show negligence, although this is difficult to demonstrate in an area of uncertain and rapidly evolving security standards and practices.[69] Even where plaintiffs can prove their case, courts generally allow damages only for concrete economic injuries associated with identity theft.[70] They refuse to recognize the other, non-economic damages that data spills create.[71] For example, some data breach plaintiffs, relying on cases that award damages for fear of illness from exposure to pathogens, have sought damages for fear of identity theft based on release of personal information;[72] courts have rejected these claims.[73] They have similarly denied claims for emotional distress damages based solely on the disclosure of a plaintiff's personal information, allowing these claims to go forward only where the plaintiff alleges fraudulent use of the information.[74] Seen in combination, these doctrines suggest that—perhaps in an effort to bolster the information economy—the law favors the collectors and users of personal information over those who suffer damage from data spills. This is similar to the law governing oil spills prior to the 1970 Clean Water Act. Highly damaging spills such as the wreck of the Torrey Canyon, or the Exxon Valdez, spurred Congress to conduct a long-overdue revision of liability

---

68. PERCIVAL, *supra* note 45, at 137.

69. Danielle Keats Citron, *Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age*, 80 S. CAL. L. REV. 241, 263-65, 268 (2007).

70. Timothy H. Madden, *Data Breach Class Action Litigation - A Tough Road for Plaintiffs*, 55 BOSTON B.J. 27, 29 (2011); *see, e.g.*, In re Hannaford Bros. Co., 613 F. Supp. 2d 108, 133 (D. Me. 2009), *rev'd in part*, Anderson v. Hannaford Bros. Co., 659 F. 3d 151 (1st Cir. 2011) (allowing a claim for concrete, economic damages).

71. *See Anderson*, 659 F.3d at 162.

72. Derek A. Bishop, *No Harm No Foul: Limits on Damages Awards for Individuals Subject to a Data Breach*, 4 SHIDLER J. L. COM. & TECH. 1, ¶ 12 (2008).

73. *See, e.g.*, Stollenwerk v. Tri-W. Healthcare Alliance, Civ. 03-0185PHXSRB, 2005 WL 2465906, *2-5 (D. Ariz. Sept. 6, 2005), *aff'd in part, rev'd in part, & remanded sub nom.*, Stollenwerk v. Tri-W. Health Care Alliance, 254 F. App'x 664 (9th Cir. 2007) (holding that injuries from exposure of personal information are only compensable if: 1) the exposure of sensitive personal information was significant, 2) the exposure significantly increased the risk of identity fraud, and 3) credit monitoring was necessary and effective in preventing identity fraud); Pisciotta v. Old Nat'l Bancorp, 499 F. 3d 629, 638-40 (7th Cir. 2007) (holding that damages of credit monitoring for information exposure "are not compensable as a matter of Indiana law"); Pinero v. Jackson Hewitt Tax Serv. Inc., 594 F. Supp. 2d 710, 716 (E.D. La. 2009) (holding the plaintiff could not recover damages for fear and credit monitoring, among other things, as these damages are "merely speculative"); Carolyn A. Deverich, Brian R. Strange, & David A. Holop, *Into the Breach: Plaintiffs Have Been Increasingly Successful in Gaining Injunctive Relief for Online Security Breaches*, L.A. LAW., Feb. 2012, at 27-28 (discussing case law in this area); Bishop, *supra* note 72 (same).

74. Bishop, *supra* note 72, ¶ 23.

limits and tort doctrines.[75] High-profile data spills, such as Target's recent release of 70-110 million customers' credit card and other personal information, may eventually prompt a similar legal transformation with respect to data security breaches. Why should society wait for the Big Data equivalent of the Exxon Valdez spill to require companies to internalize the full costs of their data security breaches? Big Data has arrived. The law need no longer nurture it. Rather, it should require the users of Big Data to internalize their external costs, thus making the information economy sustainable in the long term.

Employing the power of analogy, it should be possible to take the successful environmental law measures described above, adapt them for the data spill purpose, and so develop strategies to prevent data spills. For example, Congress could take a page from the Clean Water Act and pass legislation that authorizes government "clean up" of data spills (e.g., provision of credit monitoring, counseling and identity theft recovery services).[76] The agency that carries out the clean-up could then seek reimbursement from those responsible for the spill. This would reduce the collective action and free-rider problems that would otherwise inhibit private cost recovery lawsuits.

Like the Oil Pollution Act, such legislation could expand tort liability and require courts to recognize the non-economic damages that data spills create.[77] For example, Congress could expressly allow plaintiffs to seek damages for the emotional distress caused by the release of important personal information or from the risk of identity theft. Additionally, such a statute could establish strict liability for data spills, thus eliminating the need to prove a defendant's negligence.[78] Finally, just as the Oil Pollution Act requires oil transporters to design their ships in an environmentally protective way,[79] so the legislation could require information-intensive firms to utilize privacy by design.[80] If oil tankers must use double hulls,

---

75. *See supra* notes 57-68 and accompanying text.

76. *See* Water Quality Improvement Act of 1970, Pub. L. No. 91-224, § 11(c)(1), (f), 84 Stat. 91, 93 (1970) (codified as amended at 33 USC § 1321(c)(1), (f) (2012)).

77. *See* 33 U.S.C. § 2702(b)(2)(E) (2012) (providing that damages for "the loss of profits or impairment of earning capacity due to the injury, destruction, or loss of real property, personal property, or natural resources . . . shall be recoverable by any claimant.").

78. *See* David Vladeck, *Defining "Privacy" in a Big Data World, in* MIT BIG DATA INITIATIVE AT CSAIL, MEMBER WORKSHOP #2: BIG DATA PRIVACY: EXPLORING THE FUTURE ROLE OF TECHNOLOGY IN PROTECTING PRIVACY 7, 9 (last updated Sept. 2013), *available at* http://bigdata.csail.mit.edu/sites/bigdata/files/u9/MITBigDataPrivacy_WKSHP_2013_finalvWEB.pdf (calling for consideration of strict liability legal regime for data analytics privacy harms); Danielle Keats Citron, *Reservoirs of Danger: The Evolution of Public and Private Law at the Dawn of the Information Age*, 80 S. CAL. L. REV. 241, 265-67 (2007) (arguing for a strict liability standard, as opposed to negligence, for breaches of data reservoirs).

79. *See* Oil Pollution Act of 1990, § 4115(a), 104 Stat. at 517-18 (codified as amended at 46 U.S.C. § 3703a (2012)).

80. *See, e.g.*, ANN CAVOUKIAN, PRIVACY BY DESIGN: THE 7 FOUNDATIONAL PRINCIPLES (2011), *available at* www.privacybydesign.ca/content/uploads/2009/08/7foundationalprinciples.pdf (describing seven key features of privacy by design); ANN CAVOUKIAN & JEFF JONAS, PRIVACY BY DESIGN IN THE AGE OF BIG DATA 4 (2012), *available at* http://privacybydesign.ca/content/uploads/2012/06/pbd-big_data.pdf (discussing the application of privacy by design principles to Big Data and data analytics).

perhaps data security systems should have to employ two-factor authentication.[81] These are preliminary thoughts intended to illustrate how the oil-Big Data analogy can generate creative ideas about ways to address Big Data's privacy impacts.[82] Further work will be required to assess whether these ideas can be developed into full-fledged policy proposals.

IV. CLIMATE CHANGE LAW AND POLICY AS A MODEL

Even if all these measures were adopted, they would only address the first aspect of the Big Data privacy issue: data spills. They would do nothing to reduce the glass house effect, the deeper and more profound problem. The glass house effect is analogous to climate change. So the question becomes: is it possible to translate climate change law and policy into the realm of privacy, and so generate ideas on how to protect privacy in the era of Big Data?

A. Climate Change Law and Policy

The human combustion of fossil fuels, a significant contributor to climate change, lies at the very foundation of all industrial economies. This feature strongly influences climate change law and policy. It makes it difficult to employ traditional emission control requirements in order to achieve reductions. The sheer number and diversity of emissions sources defies the development and enforcement of such standards for all but the largest-emitting sectors. Moreover, carbon emissions are so huge and so pervasive that requiring a limited number of emitters to reduce their emissions by some feasible percentage will not achieve the decrease needed to prevent climate change and its associated ills.

Achieving this goal requires, not control requirements for fossil fuel-based energy technologies, but rather the large-scale implementation of alternative energy technologies.[83] Developed and emerging economies need to shift to clean energy

---

81. Professor Paul Ohm made this connection and generously shared it with me. Telephone Interview with Paul Ohm, Associate Professor, University of Colorado Law School (June 13, 2013). Two-factor authentication can be defined as

> a security process in which the user provides two means of identification, one of which is typically a physical token, such as a card, and the other of which is typically something memorized, such as a security code. In this context, the two factors involved are sometimes spoken of as *something you have* and *something you know*. A common example of two-factor authentication is a bank card: the card itself is the physical item and the personal identification number (PIN) is the data that goes with it.

Margaret Rouse, *Two-Factor Authentication*, TECHTARGET, http://searchsecurity.techtarget.com/definition/two-factor-authentication (last updated Sept. 2005), (emphasis in original).

82. *See* Gavetti & Rivkin, *supra* note 13, at 6 (explaining that analogies "can spark breakthrough thinking").

83. *See* EXEC. OFFICE OF THE PRESIDENT, PRESIDENT'S COUNCIL OF ADVISORS ON SCI. AND TECH., REPORT TO THE PRESIDENT ON ACCELERATING THE PACE OF CHANGE IN ENERGY TECHNOLOGIES THROUGH AN INTEGRATED FEDERAL ENERGY POLICY 8-9 (Nov. 2010), *available at* http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-energy-tech-report.pdf [hereinafter REPORT TO THE PRESIDENT ON ENERGY TECHNOLOGIES] (calling for an energy innovation agenda in

sources such as solar, hydrogen, geothermal, wind, water, and (if it can be done more safely) nuclear energy technologies.[84] Climate change policy seeks to facilitate and bring about this technological transformation. Analysts suggest that such a shift will not only reduce carbon emissions, but that it could also generate new businesses, industries, and jobs as American engineers and manufactures meet the need for clean energy technologies at home and abroad.[85] Some see this as one of the nation's most important opportunities for future economic growth and job creation.[86] The President's Council of Advisors on Science and Technology concluded that "American economic competitiveness, environmental stewardship, and enhanced security depend on picking up the pace of energy technology innovation . . . ."[87]

B. Policies to Promote the Clean Energy Transformation

Given these economic opportunities, one might assume that the market alone would meet the need for clean energy technologies. That is not the case. Three market failures—negative externalities, the public goods problem, and positive spillovers from basic research—lead the private sector to underinvest in clean energy technologies.[88] Government can correct for these market failures and so produce something closer to an optimal level of investment.[89]

A negative externality exists when the producer of a good does not have to bear the full costs of production and is, instead, able to "externalize" these costs onto others.[90] Environmental damage is an important type of negative externality.[91]

---

order to achieve a "low-carbon transformation for purposes of both climate change risk mitigation and energy security").

84. *See id.* at 3-4 (describing an integrated innovation ecosystem that includes invention, translation to commercial product adoption, and diffusion of new clean energy technologies).

85. *See* PEW CHARITABLE TRUSTS, INNOVATE, MANUFACTURE, COMPETE: A CLEAN ENERGY ACTION PLAN 10 (2012), *available at* http://www.pewenvironment.org/uploadedFiles/PEG/Newsroom/Press_Release/Innovate,%20Manufacture,%20Compete.pdf (stating that "clean energy helps create employment, manufacturing, and export opportunities while reducing the pollution and risks associated with current energy patterns and technologies").

86. *See id.* at 16 (stating, "research shows that investments in clean energy have yielded more than three times the number of jobs as have comparable investments in conventional fossil fuels"); AM. ENERGY INNOVATION COUNCIL, CATALYZING AMERICAN INGENUITY: THE ROLE OF GOVERNMENT IN ENERGY INNOVATION 9 (2011), *available at* http://americanenergyinnovation.org/wp-content/uploads/2012/04/AEIC_Catalyzing_Ingenuity_2011.pdf ("[T]o break out of our current economic malaise, America needs to innovate, manufacture and build new technologies. This is true in many sectors of our economy, and is certainly true in energy.").

87. REPORT TO THE PRESIDENT ON ENERGY TECHNOLOGIES*, supra* note 83, at 1-2.

88. *See* CONG. BUDGET OFFICE, FEDERAL FINANCIAL SUPPORT FOR THE DEVELOPMENT AND PRODUCTION OF FUELS AND ENERGY TECHNOLOGIES 1, 8-9 (Mar. 2012), *available at* http://www.cbo.gov/sites/default/files/cbofiles/attachments/03-06-FuelsandEnergy_Brief.pdf (describing externality and positive spillover issues); AM. ENERGY INNOVATION COUNCIL, *supra* note 86, at 10-11 (describing public good and positive spillover issues).

89. *Id.* at 10-11 (making the case for an active government role in energy innovation).

90. *See* ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 44 (5th ed. 2007) (explaining that "[t]he reason the market fails in the presence of external costs is that the generator of the externality does not have to pay for harming others").

91. *Id.*

When a company pollutes, it does not itself experience most of the health and environmental impacts of that pollution. Rather, it externalizes these costs onto the surrounding members of the public.[92] Since the company does not bear the majority of these costs it has little reason to invest in technologies that would reduce or prevent them, even where the benefits from such technologies (in terms of health and environmental gains) would exceed their cost. Negative externalities thus lead to sub-optimal investment in pollution reduction technologies.[93] The climate change area follows this general pattern. The operators of coal-fired power plants and other companies that burn fossil fuels and emit greenhouse gases contribute to climate change. However, much of the damage that climate change causes—the rising sea levels, spread of disease, species extinctions and droughts—does not affect these companies directly. Rather, the companies are able to externalize these costs onto the public.[94] They therefore tend to invest too little in technologies that would prevent them.[95]

The public goods problem is related, though distinct. It looks at the hypothetical company that does decide to invest in and implement clean energy technology. Assume that such a company decreases its greenhouse gas emissions and so reduces the harmful effects of climate change. While it has created a good (a more stable climate) that many might be willing to pay for, it will not be able to charge for this good because it will not be able to exclude anyone from receiving the benefits of it.[96] Climate stability is a "public good" in the case that, if it is available to one, then it is available to all.[97] Since the company cannot charge for this good, it has little incentive to invest in producing it, even where the value of the good would exceed the costs of production.[98] Thus the public goods problem, much like the negative externality one, leads to underinvestment in clean energy technology.

The third market failure is rooted in the fact that investments in basic research frequently yield general scientific knowledge that, while highly useful, cannot be the subject of a patent.[99] As a result, those who make such discoveries must share them with others, including their competitors.[100] The benefits, in effect, "spill over" from the one that created them to others who will also find them useful. Since the creator cannot exclude others from its discoveries, it cannot charge for them. This greatly reduces its incentive to engage in basic research, even where the

---

92. JAMES SALZMAN & BARTON H. THOMPSON, JR., ENVIRONMENTAL LAW AND POLICY 20-22 (3rd ed. 2010).

93. *See* CONG. BUDGET OFFICE, *supra* note 88, at 8-9.

94. *Id.*

95. *Id.*

96. PETER S. MENELL & RICHARD B. STEWART, ENVIRONMENTAL LAW AND POLICY 54 (1994) (defining "public goods" as "commodities that cannot be supplied to a given individual without at the same time enabling large numbers of other individuals to enjoy them simply because it is impracticable to exclude those other individuals from such enjoyment.").

97. *Id.* at 55.

98. *Id.*; AM. ENERGY INNOVATION COUNCIL, *supra* note 86, at 10.

99. *See* CONG. BUDGET OFFICE, *supra* note 88, at 9 (stating that "basic research . . . can create general scientific knowledge that cannot be subject to patents . . .").

100. AM. ENERGY INNOVATION COUNCIL, *supra* note 86, at 10.

benefits of doing so would exceed the costs.[101] The result is underinvestment in the fundamental research on which many important innovations are grounded.[102]

Federal clean energy policies seek to correct for these market failures. Some fund or otherwise seek to stimulate basic research.[103] Others seek to address the externality and public goods problems by requiring, or subsidizing, private sector investment in renewable energy so as to produce a more optimal level of it.[104] For the most part, the private sector supports government taking this role.[105] Federal clean energy policies include:

- Direct investments in clean energy technologies.[106]
- Loan programs[107] (including both direct loans [108] and loan guarantees) that support private sector investment in clean energy.[109]
- Tax preferences for firms that utilize or invest in clean energy, including special deductions, special tax rates, tax credits, and grants in lieu of tax credits.[110]
- Government procurement of renewable energy to meet its own energy needs.[111]

---

101. *Id.* at 16 (suggesting that"[p]rivate companies are deterred from basic research . . . because they . . . can't prevent their competitors from also capturing some of the commercially valuable knowledge gained through these investments.").

102. *See* CONG. BUDGET OFFICE, *supra* note 88 at 1 ("[U]nless the government intervenes, the amount of research and development (R&D) that the private sector undertakes is likely to be inefficiently low from society's perspective because firms cannot easily capture the 'spillover benefits' that result from it. That is particularly true at the early stages of developing a technology."); AM. ENERGY INNOVATION COUNCIL, *supra* note 86, at 10 (explaining that "the privacy sector has tended to systematically *under*-invest in R&D relative to the potential gains to society–even where a market for the desired technology exists" (emphasis in original)).

103. *See id.* at 16. Such policies assume that private companies will pick up on the results of this research and use it to develop their own patentable and marketable inventions and products. This strategy has proven effective in other areas. For example, government scientists mapped out natural resources, surveyed routes for railroads, conducted basic research on nuclear technologies and, more recently, created a distributed network of computers (ARPANET) that laid the foundation for the Internet. *Id*. at 11-12.

104. CONG. BUDGET OFFICE, *supra* note 88, at 9. The government can also achieve this by taxing the emission of greenhouse gases or other pollutants so as to internalize the environmental costs and so give firms an incentive to reduce this pollution. *Id*. This strategy, while perhaps the most cost-effective one as applied to the clean energy area, does not translate as easily into a policy for addressing the glass house effect. Accordingly, this Article focuses on subsidies in their various forms as a way of encouraging technological innovation, rather than on fees or taxes.

105. PEW CHARITABLE TRUSTS, *supra* note 85 at 32.

106. CONG. BUDGET OFFICE, *supra* note 88, at 7. One of the most successful such programs has been the Department of Energy's Advanced Research Projects Agency–Energy (ARPA-E) initiative which funds "high-risk, high-payoff research" in clean energy technologies. *Id.*

107. *Id.* at 7; EXEC. OFFICE OF THE PRESIDENT, THE PRESIDENT'S CLIMATE ACTION PLAN 7 (June 2013), *available at* http://www.whitehouse.gov/sites/default/files/image/president27sclimateactionplan.pdf [hereinafter PRESIDENT'S CLIMATE ACTION PLAN] (describing a Department of Energy loan guarantee program, which supports "investments in innovative technologies").

108. CONG. BUDGET OFFICE, *supra* note 88, at 7.

109. *Id.*

110. *Id.* at 2.

111. PRESIDENT'S CLIMATE ACTION PLAN, *supra* note 107, at 7-8. The federal government is nation's largest consumer of energy. *See id.* at 7. By insisting that its suppliers meet some of its energy

- Comprehensive planning and coordination of government and private sector clean energy strategies.[112] To this end, the President has initiated a Quadrennial Energy Review to "identify the threats, risks, and opportunities for U.S. energy and climate security, enabling the federal government to translate policy goals into a set of analytically based, clearly articulated, sequenced and integrated actions, and proposed investments over a four-year planning horizon."[113]
- Performance-based clean energy requirements such as fuel efficiency standards for the auto industry,[114] or renewable portfolio standards (generally passed at the state level), which require utilities to obtain a certain percentage of their energy from renewable sources.[115]

While researchers have not yet evaluated all of these initiatives, initial studies suggest that government programs that fund energy research and development "often yielded benefits greater than its costs,"[116] but that government funding for "later stages in the [technology] development process has been far less cost-effective."[117]

C. Promoting Innovation in Clean Data Technology

In much the same way that oil and other fossil fuels lie at the foundation of the smokestack economy, so Big Data is becoming a foundational building block of the information economy.[118] It is now so pervasive and powerful that traditional regulatory measures such as notice and choice, or purpose limitations, simply cannot constrain its privacy impacts. Notice fails because the increasingly ubiquitous collection and use of personal information renders individual notice virtually impossible, and because data analysts often cannot predict (and so cannot provide notice of) how they will use data sets.[119] Without notice, choice fails as

---

needs through renewable energy sources, the government can help to create a market for early renewable technologies. This can be especially helpful at the early stages of deployment while producers are working to bring down the cost of these technologies.

112. *Id.*

113. This is needed to pull together the "amalgam" of energy-related policies and meld them into an integrated approach that can utilize resources more efficiently and effectively. *See* PRESIDENT'S CLIMATE ACTION PLAN, *supra* note 107, at 8.

114. *Id.*

115. *See generally* Ivan Gold & Nidhi Thakar, *A Survey of State Renewable Portfolio Standards: Square Pegs for Round Climate Change Holes?* 35 WM. & MARY ENVTL. L. & POL'Y REV. 183 (2010) (surveying state renewable portfolio standard programs).

116. *See* CONG. BUDGET OFFICE, *supra* note 88, at 9 (citing NAT'L RESEARCH COUNCIL, ENERGY RESEARCH AT DOE: WAS IT WORTH IT? ENERGY EFFICIENCY AND FOSSIL ENERGY RESEARCH 1978 TO 2000 (Nat'l Acad. Press 2001)).

117. *Id.* at 2. *See also* REPORT TO THE PRESIDENT ON ENERGY TECHNOLOGIES, *supra* note 83, at 3-4 (dividing the innovation process into four phases—invention, translation, adoption and diffusion—and arguing that federal support is more effective when targeted at the first of these stages).

118. *See* MAYER-SCHÖNBERGER & CUKIER, *supra* note 2, at 16 (explaining that Big Data "is becoming a significant corporate asset, a vital economic input, and the foundation of new business models. It is the oil of the information economy.").

119. Ira S. Rubinstein, *Big Data: The End of Privacy or a New Beginning?*, 3 INT'L DATA PRIVACY L. 74, 78 (2013); WORLD ECONOMIC FORUM, *supra* note 26, at 11.

well. Individuals cannot give meaningful consent to uses of which they have had inadequate notice.[120] Purpose limitations create a direct conflict with Big Data, the value of which lies in continuously re-purposing massive data sets and finding new and unanticipated uses for them.[121] For all of these reasons, traditional privacy regulation—notice, choice, purpose limitation—will prove ineffective as applied to Big Data. As in the climate change area, the solution lies more in technological transformation than in traditional regulatory responses.

Mitigation of the glass house effect will require the development and use of new "clean data" technologies[122] that allow society to derive the benefits of Big Data, while minimizing its privacy impacts. Such technologies are already beginning to emerge.[123] Privacy-technology experts, Oliveira and Zaïne, have defined a set of criteria for "privacy-preserving data mining technologies" (PPDM)—data mining technologies that "encompass[] the dual goal of meeting privacy requirements and providing valid data mining results."[124] A number of emerging technologies seek to achieve this goal. For example, differential privacy techniques allow analysts to query Big Data but introduce a degree of "noise" into the response.[125] This "noise" creates an interference that is large enough to disguise the presence or absence of any individual in the data set, yet small enough to ensure that the answer provided is still useful.[126] On another front, efforts are being made to develop "new techniques to securely de-identify data."[127] Privacy-

120. Rubinstein, *supra* note 119, at 75; Tene & Polonetsky, *supra* note 7, at 260-61.

121. MAYER-SCHÖNBERGER & CUKIER, *supra* note 2, at 2, 103; WORLD ECONOMIC FORUM, *supra* note 26, at 11.

122. *See* CAVOUKIAN & JONAS, *supra* note 3, at 8 (stating that attention should "shift from compliance with FIPs to proactively embedding privacy into the design of new technologies").

123. *See, e.g.*, Klarreich & Simons Science News, *supra* note 24 (discussing technologies that allow us to "get at these data without revealing private information"). *See also* CYNTHIA DWORK, MICROSOFT RESEARCH, DIFFERENTIAL PRIVACY 8-11 (2006), *available at* http://pdf.aminer.org/000/267/828/differential_privacy.pdf (describing differential privacy techniques that allow users to ask questions of large data sets while preserving individual privacy); Stanley R. M. Oliveira, *Privacy-Preserving Data Mining*, in ENCYCLOPEDIA OF DATA WAREHOUSING AND MINING (2d ed. 2008) (describing emerging data mining technologies that protect privacy and defining the groundwork for further research and development in this area).

124. OLIVEIRA & ZAÏANE, *supra* note 9, at 3.

125. *See* Klarreich & Simons Science News, *supra* note 24 (describing how a differential algorithm might "add noise" to an answer by adding or subtracting some number before returning the answer). *See also* DWORK, *supra* note 123 at 9-11 (giving examples of how noise may be added in certain formulas).

126. *See* DWORK, *supra* note 123, at 8-9. *See also* MICROSOFT CORP., DIFFERENTIAL PRIVACY FOR EVERYONE 3-5 (2012), *available at* http://www.microsoft.com/en-us/download/details.aspx?id=35409 (showing an example of how differential privacy can yield useful and privacy-preserving answers). "A differentially private data release algorithm allows researchers to ask practically any question about a database of sensitive information and provides answers that have been 'blurred' so that they reveal virtually nothing about any individual's data–not even whether the individual was in the database in the first place." Klarreich & Simons Science News, *supra* note 24. To achieve this blurring effect, differential privacy establishes a piece of intermediary software that stands between the one asking the questions and the database itself. MICROSOFT CORP., *supra* at 3. This "privacy guard" calculates the minimum amount of noise needed to protect individual identities, and adds it to the answer that it provides to the analyst. *Id*.

127. CASTRO, *supra* note 25, at 2.

preserving data mining and successful de-identification (assuming that the latter is possible)[128] offer a glimpse of a more optimistic future. However, most of these technologies remain at the experimental stage and have not yet been developed as commercial applications.[129]

The further development of privacy-preserving data mining technologies will likely confront the same market failures that the generation of clean energy technologies has encountered. In the clean data context, as in the clean energy one, investments in basic research will produce positive spillover effects that private investors cannot capture and monetize.[130] As a result, the private sector will provide a sub-optimal level of investment in the development of such technologies. Moreover, a company's privacy impacts constitute a negative externality, while its investments in privacy protection create a public good—consumer trust in the information economy as a whole. These market failures, too, will produce sub-optimal investment in privacy-preserving data mining technologies. The rationale for government support for clean energy technologies thus applies with equal force to the development of these "clean data" technologies. The United States should develop a privacy-preserving data mining technology agenda[131] that parallels the nation's strategy for promoting clean energy.[132]

Drawing from the clean energy playbook, such an agenda could include:

- Direct investments to support the development of technologies and techniques, such as differential privacy, that allow analysts to utilize Big Data but reduce the privacy impacts that result from such activities. These investments should focus on support of basic research into such technologies. It should further seek to support the progress of differential privacy and other existing techniques in order to move them past the experimental stage and bring them to the point of commercialization.
- Loan programs (including direct loans and loan guarantees) that support private sector investment in privacy protective data mining technologies.
- Tax preferences for data analytics or other firms that utilize or invest

128. *See generally* Paul Ohm, *Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization*, 57 UCLA L. REV. 1701 (2010) (arguing that efforts to de-identify have failed and regulators must respond).

129. *See* MICROSOFT CORP., *supra* note 126, at 6 (explaining that differential privacy, while promising, "is [sic] still a research-level technology, not a commercial product, and . . . its potential implementation in real-life research and commercial scenarios . . . will present mathematical, computational, and policy challenges that will need to be addressed before it can go into production."); MARTIN MEINTS & JAN MÖLLER, PRIVACY PRESERVING DATA MINING: A PROCESS CENTRIC VIEW FROM A EUROPEAN PROSPECTIVE 12 (2007) (stating that "PPDM is still an area of research and not readily implemented on the market yet.")

130. *See supra* notes 88-89 and accompanying text (explaining that basic research creates positive spillovers).

131. *See* CASTRO*, supra* note 25, at 2 (stating that "the U.S. government should create and fund a research and development (R&D) roadmap for privacy.").

132. *See generally* REPORT TO THE PRESIDENT ON ENERGY TECHNOLOGIES, *supra* note 83 (setting out a strategy for promoting clean energy).

in such technologies. These could include special deductions, tax rates, tax credits, and grants in lieu of tax credits.

- A preference for firms that employ privacy-preserving data mining technologies in the federal government's own purchase of data analytics services. The federal government is a large consumer of these services. If it gave a preference to such firms this could jump-start the market for privacy-enhancing data mining technologies and so promote the development and commercialization of such products.
- Planning and coordination of government and private sector strategies for the development of privacy-preserving data mining technologies. The federal government could play a coordinating role in which it sees the big picture, identifies funding gaps or particularly promising areas of development, and seeks to direct resources towards these priority areas. As with the Federal Quadrennial Energy Review,[133] this should be a comprehensive planning process that involves relevant agencies and stakeholders and seeks to integrate existing efforts in order to marshal resources in the most effective way possible

Such policies, and the further development of PPDM technologies, are essential for the preservation of privacy in a Big Data world. They will prove useful to government agencies and private businesses that apply data mining to personal information.[134] On a broader level, they make sense for any society that wishes to make greater use of Big Data and enjoy the many benefits it can provide. As mentioned above, a lack of investment in such technologies, and a consequent failure to address Big Data's privacy impacts, could cause a social and political backlash that could impede further growth in this area.[135] Privacy protections are essential to "unlocking the value of personal data."[136] The development of PPDM technologies may also provide an important economic opportunity. Many countries and businesses will need such technologies if they are to exploit fully the benefits of Big Data and data analytics. Those who develop and patent these new technologies will possess a competitive advantage in a global market for such goods and services.

> Additional development of privacy tools could also have a positive economic impact. Investments in developing technological solutions to privacy problems would help create a network of developers with expertise in this domain. Developers of such tools would likely be even more competitive in countries with strict privacy regulations, where there may be a stronger market for privacy-

133. PRESIDENT'S CLIMATE ACTION PLAN, *supra* note 107, at 7-8.

134. *See* CASTRO, *supra* note 25, at 2 (stating that "every government agency that uses personally identifiable information (PII) might benefit, either directly or indirectly, from advances in privacy-preserving data mining or new techniques to securely de-identify data. Similarly, industries such as health care and financial services would benefit from this research as well.").

135. *See supra* note 23 and accompanying text.

136. *See supra* note 26 and accompanying text.

enhanced products and services.[137]

Politicians and policymakers in the U.S. have made the case that government investment in clean energy technologies could pay off in increased business activity, exports and jobs.[138] Investments in clean data technologies could too, though probably on a smaller scale.

The federal government does not yet appear to have fully recognized the important role that privacy-preserving data mining technologies can play. While it has put been working to develop laws that will protect privacy better, and has been enforcing existing rules, it has put "relatively little effort . . . into considering how new technology might address many of the same privacy challenges."[139] It should now turn its attention to this policy challenge, just as it has done in the area of clean energy.

V. CONCLUSION

"Big Data is the new oil" is a powerful analogy, indeed. In six short words it helps us to envision Big Data's great potential future. It also helps us to grasp—intuitively, and immediately—the threats that it poses.[140] Data spills *are* like oil spills; they damage individual lives and livelihoods just like oil spills do. The contemporary threat to individual privacy—what this Article has called the "glass house effect"—*is* like climate change. We hear about the ways that corporations and the NSA are collecting, aggregating, and mining our data. The ads we receive on our computers or mobile devices give us inklings as to what is happening out there in the data universe and how it is affecting us. Yet the shift is so profound and pervasive that it almost escapes comprehension. We know what is happening, and yet we do not know it. We need a framework, a concept, an analogy to help us grasp something that cannot be understood in its entirety.

Climate change is that analogy. The change is in the very air around us; it is in the social climate in which we live. A world that once offered some protection for our private comings and goings, our intellectual or political interests, our eccentricities, maladies and vulnerabilities, increasingly records them and makes them visible. The great paradox is that, just as digitization and the Internet open wide doors for individual exploration, expression and growth, so they increasingly turn us into a surveillance society in which we feel observed and pressured to conform. It is like industrial society, which provides us with new goods and tools that enhance our lives and at the same time threatens to despoil the natural environment on which we depend.

The "Big Data is the new oil" analogy is valuable because it can help us to see both the tremendous promise of, and the threats from, Big Data. It can enable us to grasp, in an instant, something big and mysterious, wonderful and dark, that is

---

137. CASTRO*, supra* note 25, at 2.

138. *See* PEW CHARITABLE TRUSTS, *supra* note 85, at 10; AM. ENERGY INNOVATION COUNCIL, *supra* note 86, at 9; REPORT TO THE PRESIDENT ON ENERGY TECHNOLOGIES*, supra* note 83, at 1-2.

139. *Id.* at 1.

140. *Cf.* Gavetti & Rivkin, *supra* note 13, at 2 (explaining that "[a]nalogical reasoning makes enormously efficient use of the information and mental processing power that strategy makers have.").

happening in the world around us. It can also help us to generate solutions.[141] Environmental law has developed measures to reduce oil spills and to prevent climate change. These steps are incomplete and imperfect, particularly where climate change is concerned. Yet they reflect a body of effort and thinking that, over time, has crystalized into specific policies, legal doctrines, and government investments. We need not start from scratch in thinking about how to protect privacy in the era of Big Data. We can take advantage of the work that has already been done in the environmental field, translate it to the privacy issue, and emerge with a starting point for the new task. That is the power of analogy; we should use that power.

---

141. *Cf. id.* at 3 (describing how analogical thinking can "spark creativity").