The Glass House Effect: Why Big Data is the New Oil, and What to Do About It.
Professor Dennis Hirsch[1]

"Data is the new oil," Clive Humbly announced in 2006.[2] Just recently, IBM CEO Virginia Rometty updated the phrase, explaining that "big data" is the new oil.[3] The analogy resonates. Data flows like oil. One must "drill down" into data to extract value from it. Data is an essential resource that powers the information economy in much the way that oil has fueled the industrial economy. Data promises a plethora of new uses—diagnosis of diseases, direction of traffic patterns, etc. – just as oil has produced useful plastics, petrochemicals, lubricants, gasoline, and home heating. "Big data is the new oil" has not only become a popular meme; it is a banner behind which we can march, an optimistic declaration of the way forward.

But the comparisons frequently ignore oil's negative side. Tankers run aground and spill their deadly black cargo. The Deepwater Horizon platform collapses in flames and raw oil gushes into the Gulf for weeks. This too must be included in the analogy. Data spills occur with the regularity of oil spills. The victim of identity theft, bogged down in unwanted credit cards and bills, is just as trapped and unable to fly as the bird caught in the oil slick, its wings coated with a glossy substance from which it struggles to free itself.


Chronicle / Frederic Larson

As the data sets get bigger the threat, too, grows. Big data is like a massive oil tanker navigating the shoals of computer-savvy criminals and human error. Yes, big data make us smarter and wealthier and our lives better. But that dream has a dark, viscous underside that threatens to pollute the information ecosystem.

How to proceed? Environmental law reduces oil pollution without undermining the fossil fuel-based economy. Can we look to it for strategies that will allow us to reap big data's many benefits, while reducing its negative impacts?

---

[1] 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.

[2] Clive Humbly, ANA Senior marketer's summit, Kellogg School (2006); *see* Michael Palmer, *Data is the New Oil,* blog post available at http://ana.blogs.com/maestros/2006/11/data_is_the_new_oil.

[3] http://siliconangle.com/blog/2013/03/11/ibms-ceo-says-big-data-is-like-oil-enterprises-need-help-extracting-the-value/

The history of oil pollution law is highly instructive. In the 19th Century, judges and legislators shaped the law to encourage the production and transportation of oil. Maritime tort law recognized property damage from oil spills, but not injuries to fishing, tourism and other such affected industries. Traditional tort doctrines required plaintiffs to show negligence—a difficult task in a risky field where even the careful could spill their cargo. Collective action and free rider problems further reduced the incentives to bring such a suit since, when many suffer a small harm, no single person has the incentive to bring the suit or to organize the group to do so. Finally, as if tort liability were not yet sufficiently constrained, Congress passed the Limited Liability Act of 1851 which capped oil spill damages at the value of the vessel and freight remaining after the accident.[4] This statute, whose original purpose was to facilitate the transportation of otherwise uninsurable cargo, came to produce absurd results. 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, resulted in only $50 in damages—the value of the sole remaining lifeboat.[5] Clearly, something needed to be done.

Congress gave its answer in the 1970 Clean Water Act[6] and, responding to the massive Exxon Valdez oil spill, the 1990 Oil Pollution Act.[7] Together, these statutes re-write oil pollution law. They allow the government to clean up an oil spill and then bring an action against the responsible party to recoup the clean-up costs.[8] This overcomes the collective action and free rider problems that undermine private tort actions. The Oil Pollution Act recognizes new causes of action for damage to economic, as opposed to property, interests.[9] The statutes provide for strict liability, thereby relieving plaintiffs of the difficult task of demonstrating negligence. They greatly increase the liability limits.[10] Finally, the Oil Pollution Act requires all new oil transportation vessels operating in U.S. waters to employ double hull technology that greatly reduces the chance of an oil spill.[11] The statutory scheme has reduced spills by oil-transporting vessels.

This environmental law success story offers important lessons for big data. Like the early laws governing the oil industry, today's doctrines appear designed to encourage the production and transfer of the "new oil." Following a data spill, courts generally allow damages only for the concrete economic injuries associated with identity theft. They refuse to recognize the other, non-economic damages that data spills create – the increased risk of identity theft and the anxiety that that risk produces; the sense of violation and exposure that comes from release of one's personal data.[12] As in the oil pollution context,

---

[4] 46 U.S.C. 183 (1988).

[5] Jeffrey D. Morgan, *The Oil Pollution Act of 1990: A Look at its Impact on the Oil Industry,* 6 FORD. ENV. L. J. 1, 2 (1994)

[6] Pub. L. No. 91-224, 84 Stat. 91 (1970) (codified as amended in scattered sections of 33 U.S.C.)

[7] Pub. L. No. 101-380, 104 Stat. 484 (1990).

[8] Water Quality Improvement Act, § 11(c)(1), 84 Stat. at 93.

[9] Oil Pollution Act § 1002(b)(2), 104 Stat. at 490; *see also* Kenneth M. Murchison, *Liability Under the Oil Pollution Act: Current Law and Needed Revisions*, 71 LA. L. REV. 917 (2011).

[10] Oil Pollution Act, § 1004(a)(1), 104 Stat. at 491-92. The original Oil Pollution Act raised the limits to the greater of $1200 per ton, or $10,000,000 for a vessel greater than 3000 tons, or $2,000,000 for a smaller vessel. *Id.*

[11] 46 U.S.C. 3703(a) (1988 & Supp. IV 1992).

[12] *See, e.g., Pinero v. Jackson Hewitt Tax Service, Inc.*, No. 08-3535, 2009 U.S. Dist. LEXIS 660, (E.D. La. January 7, 2009); Kelley, Drye & Warren, Consumer Finance Law Blog, *Fears of Future Identity Theft Generally Not Sufficient to Establish Actual Damages in a Lawsuit,* available at http://www.consumerfinancelawblog.com/2009/03/articles/privacy/fears-of-future-identity-theft-generally-not-sufficient-to-establish-actual-damages-in-a-lawsuit/ (last visited June 30, 2013).

negligence is difficult to prove in the complex area of data security. Collective action and free-rider problems abound. Why should any individual bring the suit that requires a company to provide increased data security for *all* its customers? Data breach notification statutes require firms to bear the cost of providing notice to affected persons, but not the full cost of the injuries that their breach has caused. While these laws provide a notice and deterrent function that makes them far more useful than the 1851 Limited Liability Act, the liability that they create is limited. Why should we wait for the big data equivalent of the Exxon Valdez spill to change this system and require companies to internalize the full costs of their data security breaches? Big data has arrived. We no longer need to design the law to subsidize it. Rather, we need laws that require big data to internalize its externalities and so make the information economy sustainable in the long term.

Environmental law provides a possible model for doing this. As with the Clean Water Act and Oil Pollution Act, Congress could pass legislation that authorizes, and provides funding for, a government agency to clean up after data spills (e.g. to identify root causes, assess the extent of the breach, and provide credit monitoring and identity theft recovery services to consumers). The agency could then seek reimbursement from the responsible party. This would overcome the collective action and free-rider problems that would otherwise inhibit private lawsuits. Like the Oil Pollution Act, such legislation could also expand tort liability and require courts to recognize the non-economic damages that data spills create. It could establish strict liability for data spills and so eliminate the need to prove defendant's negligence. Just as the OPA requires ships to adopt an environmentally-protective design, so the legislation could require firms to adopt privacy by design. If oil tankers must use double hulls, perhaps data security systems should have to employ two-factor identification.[13]

In the environmental area, increased liability for oil transporters only gets you so far. The real solution is the development of clean energy technologies—solar, wind, hydro and geothermal power--that can substitute for oil and produce far smaller environmental impacts. The same should be true for big data. The answer is not simply to punish data spills. It is to prevent them through new, "clean data" technologies and privacy-protective business models. Recently, the United States and other countries have engaged in a massive push to develop clean energy technologies. They know that these innovations are needed, not only for preserving health and quality of life, but for future economic competitiveness. As data sets grow larger and larger, could the desire for privacy and consumer trust ramp up demand for clean data technologies? Could these innovations, too, be technologies of the future that form the basis, not only for better data security and privacy protection, but for the emergence of a "clean data" sector that makes us more competitive? Should we fund a push for innovation with respect to encryption, data anonymization and other clean data technologies?[14]

Society needs such investment. The use of oil and other fossil fuels has created environmental impacts that go far beyond oil spills. It has rapidly built up a layer of carbon in the earth's atmosphere that leads

[13] Professor Paul Ohm made this connection and generously shared it with me. Two-factor identification 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." http://searchsecurity.techtarget.com/definition/two-factor-authentication (last visited June 30, 2013).

[14] Professor Ohm suggested the importance of such a project.

to the greenhouse effect and global warming.[15] This threatens to destabilize the climate, damage ecosystems and make earth less hospitable for human and other forms of life.[16] In a similar way, new technologies and business models are building a layer of personal data. The increase is far faster that it has been with carbon. The amount of data in the world is doubling every two years -- an exponential, extraordinarily rapid rise.[17] As the recent NSA scandal has revealed, this build-up affects us. It makes us feel that our personal information is being collected and sorted. It makes us more careful about how we behave. It inhibits us from experimenting or engaging in controversial activities, from becoming the persons we want or need to be.

This is not the greenhouse effect, but the *glass house effect*. It is as if we were increasingly living in a glass house whose transparent walls allowed the hot glare of public scrutiny to reach in and scorch our most private selves. What else can we call it when companies save our search queries and web activity and communications and share them with each other, or with the government? While climate change acts on the physical world, the glass house effect acts on our inner lives. It focuses too much light on us and, in so doing, stunts the growth of the "inviolate personality"[18] which requires shade and shelter in which to flourish. Like the greenhouse effect, the glass house effect produces conditions that are less favorable to life – to a full, human life. If the evolution of big data continues on its current track we will pass on to our children a depleted ecosystem for the cultivation of the human personality. This is the threat that the tankers of big data pose as they traverse the globe.

The optimistic claim that "big data is the new oil" is indeed helpful. It both shows us the tremendous upside of this new phenomenon, and points to the threats that big data, like oil, poses. It should motivate us to find sustainable ways to utilize this highly valuable new resource—methods that allow us to enjoy the benefits of big data, while preserving fertile ground for personal development.

---

[15] *See e.g.,* International Panel on Climate Change, Fourth Assessment Report: Climate Change (2007) (describing scientific findings on climate change).
[16] *Id.*
[17] Steve Lohr, *The Age of Big Data*, N.Y. TIMES (Feb. 11m, 2012), available at http://www.nytimes.com/2012/02/12/sunday-review/big-datas-impact-in-the-world.html?pagewanted=all (last visited June 30, 2013).
[18] Louis Brandeis & Samuel Warren, *The Right to Privacy,* 4 HARV. L. REV. 193, 205 (1890).