Big Data Sustainability

An Environmental Management Systems Analogy

Dennis D. Hirsch*
Jonathan H. King**

Introduction

It is commonly proclaimed that "big data is the new oil."[1] This is true in the sense that data, like oil, constitutes a critical and therefore valuable resource on which our society depends. But it is also true in the sense that big data, like big oil, can generate major, if unintended, negative impacts. Where big oil produces oil spills, smog and climate change, big data can lead to data spills, privacy violations,[2] identity pollution,[3] and harmful discrimination.[4] In both contexts, uses at scale

---

* Geraldine W. Howell Professor of Law, Capital University Law School.

** LL. M Graduate in Intellectual Property and Technology Law, Washington University. The views and opinions expressed by the author are not necessarily the views of his employer. For helpful comments and perspective we would like to thank Woody Buckner, Andrew Clyne, Jay Cuthrell, Tony Falco, Jason Hoffman, Matthew Johnson, Ryan Kirk, Gavin McMurdo, Jim Newkirk, Stephane Ribot, Neil M. Richards, Richard Seroter, David Shacochis, David Trowbridge, Jim Whitehurst and Jared Wray.

[1] *See* 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/.

[2] See Dennis D. Hirsch, *The Glass House Effect: Big Data, The New Oil, And The Power OF Analogy*, 66 MAINE LAW REVIEW, 374 (2014).

[3] *See* Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42-32 (2013); Neil M. Richards & Jonathan H. King, *Big Data* Ethics, 49 Wake Forest Law Review 393 (2014).

[4] *See* Dennis D. Hirsch, *That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority,* 103 KY L. J. 345 (2015).

Word Count: _____

produce not only tremendous societal benefits, but also meaningful, unintended societal externalities that can run afoul of regulators. These externalities are to the big data economy what environmental damage has been to the smokestack economy: a negative by-product of otherwise beneficial and productive business activity.

At this formative moment of mass big data adoption we can learn from environmental management practices developed to manage the negative externalities of the industrial revolution. Today, organizations globally wrestle with how to extract valuable insights from diverse data sets without invading privacy, causing discrimination, harming their brand or otherwise undermining the sustainability of their big data projects. Leaders in these organizations are thus asking: What is the right management approach for achieving big data's many benefits while minimizing its potential pitfalls? Leveraging our analogy, we propose in this paper that Environmental Management Systems ("EMS") provide a good reference model for organizations to consider for managing their expanding big data operations.

We make this case in three parts. First we show that, along with its many benefits, big data creates negative externalities that are structurally similar to environmental pollution. This suggests that management strategies to enhance environmental performance could provide a useful model for businesses seeking to improve their privacy performance. Second, we chronicle environmental management's historical progression from a back-end, siloed approach to a more collaborative and pro-active "environmental management system" approach. We argue that Internal Review Board-inspired Consumer Subject Review Boards are similar to traditional environmental management. They sit at the end of the project development process and review proposals against identified criteria. An approach modeled after environmental management systems – a Big Data Management System ("BDMS") approach – would be integrated instead of compartmentalized; preventative rather than reactive. We believe that it offers a more effective model for managing data analytics operations to prevent negative externalities. Finally, we show that the BDMS approach aligns with: A) Agile software development and Dev Ops practices which companies use to develop and maintain big data applications, B) best practices in privacy design and engineering and C) emerging trends in

organizational management theory. This makes it a natural fit for big data management.

If big data is to achieve its many transformative benefits, the businesses leading its growth need to figure out how to minimize its unwanted, negative impacts. This is the same path that environmental management has traversed from the time that companies paid little attention to their environmental impacts to the present day when many advanced companies seek to make their operations more environmentally and socially sustainable. At this critical, formative moment when organizations want to leverage big data to revolutionary ends, we can learn from environmental management systems how to move in the direction of big data sustainability.

I. Big Data Environmental Analogy

A true story helps to illustrate the challenge that data analytics companies face today. It concerns inBloom, a non-profit financed by $100 million in Gates Foundation and Carnegie Corporation funding. inBloom sought to collect student data from public school districts across the country, develop analytics-based educational recommendations for individual students, and then funnel these to classroom dashboards. Teachers would use the recommendations to provide their students with more personalized education.[5] This noble idea soon ran into problems. Parents of the schoolchildren worried that the 400 fields of data inBloom was collecting, including information on family violence, student disabilities and other topics that might cast their child in a negative light, might attach to their children as they moved through life and constrain their educational and employment opportunities.[6] They worried about who else would gain access to this data, either because inBloom intentionally shared it with them or because it suffered from a data

[5] *See* Elizabeth Dwoskin & Lisa Fleisher, *Parental Opposition Fells inBloom Education-Software Firm: Privacy Concerns Over Student Data Lead Lead Company to Close*, WALL STREET JOURNAL (April 21, 2014), *available at* http://www.wsj.com/articles/SB10001424052702304049904579516111954826916 (describing how parental opposition caused schools to withhold student data and so forced inBloom to close).

[6] *Id.*

security breach. Parents began to protest inBloom's collection and use of of data about their children. School districts, and then entire states, refused to share student information with inBloom. Deprived of the data that it needed to operate, this promising, well-intentioned initiative shut itself down.

The inBloom story, and the business difficulties that it illustrates, bear a strong resemblance to the challenges that smokestack industries have faced on the environmental front. Here, too, beneficial business activities create significant externalities (real or perceived) that engender public opposition and become a constraint on further industrial development. This symposium is focused on Consumer Subject Review Boards and whether they can provide a useful model for data analytics management. But it is really posing a bigger question: what management approach should data analytics businesses employ to minimize their externalities and so achieve the tremendous benefits of which they are capable? Environmental management has a lot to tell us about how to answer this question.

Companies have made substantial progress with environmental management. Some of the same firms that once polluted with abandon now prioritize environmental compliance and have adopted sustainability as part of their core mission. Others have gone beyond this and found ways to turn environmental performance into competitive advantage by making environmentally-friendly products or building trust in brand. The transition is far from complete and much more work remains to be done. But it has already generated valuable strategies for maximizing production benefits and minimizing externalities. The growing data analytics sector can benefit from these lessons.

It is hard for organizational leaders to see clearly through the hype of big data, let alone properly manage potential negative externalities. It is unclear whether any given project may have more risks than benefits or that big data concerns are just a more complicated "Y2K" over exaggeration. Using the environmental analogy helps technical and non-technical, legal and non-legal, commercial and governmental leaders come to grips with how best to realize benefits of big data analytics while managing potential negative externalities. We can add a measure of predictability by learning from environmental policy, regulation and management precedents from the industrial revolution. Environmental management systems provide a particularly useful model.

II. Environmental Management Systems

Environmental management has climbed a steep learning curve. In the early days of environmental compliance, companies placed their environmental managers at the end of the planning process. The design and production departments would decide what they wanted to make and how they would produce it. Then, after completing much of their planning work, they would consult the environmental manager to find out what they needed to do to comply with environmental laws. All too often, the environmental manager ended up telling the business teams what they could *not* do and sending them back to rework their plans. Design and operations professionals came to view environmental managers as internal cops and environmental management as a necessary evil.[7]

This type of back-end environmental management strengthened compliance by the book but hurt production and stifled innovation in environmental compliance itself. Environmental managers convince product groups to fix mistakes and oversights in order to meet legal requirements. This took more time and, when it required that plans be changed, imposed major delays. Moreover, the solutions themselves often took the form of end-of-pipe pollution control technologies bolted on at the final stage of the production process. Since most of the product and process design planning had occurred long before the environmental manager got involved, companies missed opportunities for upstream solutions (e.g. choices about product design or raw materials) that could have prevented the pollution from being created in the first place and addressed the issue at far lower cost.

While environmental management continued on this course, the broader management of industrial production began to change. Catalyzed by the work of statistician W. Edwards Deming, new production approaches emerged in the auto industry. Deming examined traditional methods for ensuring quality at the big US auto companies where the production line never stopped and employees identified and fixed defects at the end of the line. Deming advocated improving quality by optimizing the manufacturing system as a whole so that it did not produce defects,

[7] See Dennis Hirsch, How To Improve Privacy Protection by Adpating and Using Environmental Management Tools, Privacy Officers Advisor, Aug. 2005, Vol 5, No 11.

rather than by fixing defects at the end of the line. The American auto makers, at the height of their power, resisted Deming's ideas at the time. A small automobile manufacturer in Japan named Toyota embraced them. Deming's ideas became the heart of the Toyota Production System or "TPS" that produced dramatic quality improvements while reducing costs and improving customer satisfaction. Eventually, TPS evolved in the 80s and 90s where it influenced Total Quality Management and, more recently, Lean Manufacturing.

Forward-thinking companies started to apply Deming's and Toyota's methods to environmental management. These pioneers viewed excess pollution as a type of defect. Rather than capturing a pollution defect at the end of the production process, as most environmental compliance efforts did, the system could be optimized to minimize pollution in the first place. Pollution would be prevented, rather than just controlled. The result of this application of Total Quality Management to the environmental arena was the Environmental Management System.

Environmental Management Systems (EMS) differ significantly from traditional environmental management. Instead of being siloed and cut off from others in the planning process, an EMS emphasizes an integrated approach that brings down the walls separating various business departments. Design, production and environmental managers work together to figure out how to create products and processes that cost-effectively minimize pollution, comply with environmental laws and produce quality products. Working with the design and production teams, the environmental manager becomes a collaborator and an innovator, not an internal cop. Instead of a case-by-case approach, collaborative teams look at optimizing the entire system to prevent pollution. Instead of reacting to pollution, collaborative teams innovate to prevent it from being created in the first place. Frequently, these front-end, pollution prevention solutions end up saving organizations money as compared to end-of-pipe controls. Studies of EMS's demonstrate their ability to promote pollution prevention, enhance compliance and reduce compliance costs.[8]

[8] See Christopher Sheldon & Mark Toxon, Environmental Management Systems (2006).

III. Big Data Management Systems

Viewed from the perspective of environmental management, Consumer Subject Review Boards, and the IRB's on which they are based, look a lot like the early, back-end approach. The CSRB sits at the end of the design and planning process. Much like the environmental manager of old, the CSRB would receive proposals from the business teams and evaluate them for their privacy and discriminatory effects. Where it found significant issues, it would send them back for improvement and resubmission. This would impose delays and might generate the same kind of resentment that early environmental managers experienced and that IRB's themselves have engendered in the university context.

Management of data analytics should learn from the history of environmental management. An EMS-like model—a Big Data Management System approach—would have privacy professionals, data scientists and programmers collaborating together to be aware of potential privacy and discriminatory impacts as they extract valuable insights from diverse data sets to test and develop their algorithms. A BDMS management structure would have the person responsible for mitigating privacy and discriminatory impacts present at the front end of the process as part of the agile team working any given big data project. This manager would ensure that product design, engineering and operations teams are seeing both the benefits and the potential privacy and discrimination issues as they design and implement algorithms and applications. This would reduce the need for late-stage evaluation of the product since societal implications—both beneficial, and potentially harmful—would be considered throughout. Just as EMS's help prevent pollution, so BDMS's should help prevent privacy and discriminatory impacts. Just as pollution prevention is less costly than end-of-pipe pollution controls, so prevention of privacy and discriminatory impacts from the front end should be less expensive and more streamlined than a cumbersome review process at the back end.

There is another important reason to consider an approach grounded in the EMS model: It fits naturally with the way that companies increasingly test, develop and operate their big data systems and applications. Companies have increasingly moved from top-down, compartmentalized models such as 'water-fall' to adopt Agile project

management and DevOps software development methods that embrace an emergent and collaborative approach.[9] Originating from the same Deming-inspired Lean Manufacturing roots as EMS, Agile and DevOps seek to make continuous improvements throughout the process, not at the end of it. [10] A 'minimum viable product' is conceived, launched and then rapidly iterated upon by teams of people to improve as they operate.[11] By making privacy leaders part of agile teams, privacy and discriminatory issues can become part of defining the minimum viable product at the outset and part of identifying and making privacy-related improvements as they arise.

A BDMS model is naturally aligned with Agile and DevOps mindsets. Privacy and anti-discrimination principles can be seen as an engineering restraint to continuously improve upon, not deny, evade,[12] or simply to meet. In the Phoenix Project, a leading book on Dev Ops, the importance of addressing system restraints is explored. The Phoenix Project is the code name for an important new retail application at a fictional company called Parts Unlimited. In the book, a yoda-like outside advisor named Erik is brought in by the board to help the newly appointed VP of IT recover from a series of IT outages, security breaches and delays in launching the all important Phoenix Project. One of the first lessons Erik teaches the VP of IT is that failure to address restraints causes unplanned work which breaks sustainable operations:

> *"Your job as VP of IT Operations is to ensure fast, predictable and uninterrupted flow of planned work that delivers value to the business while minimizing the impact of unplanned work, so you can provide stable, predictable and secure IT service."*[13]

[9] See Agile Manifesto – 12 principles and commentary.

[10] See Dan Woods, Why Lean and Agile Go Together, Forbes, Jan. 12, 2010 available at http://www.forbes.com/2010/01/11/software-lean-manufacturing-technology-cio-network-agile.html. See also, Poppiendick, The Lean Mindset

[11] See Lean Startup, origination of term "MVP".

[12] See William Boston, Volkswagon Shares Dive on Emission Woes, The Wall Street Jounral (Nov. 4, 2015).

[13] The Phoenix Project, Kindle 1340 of 5829

In a post Snowden era with no more Safe Harbor, we have clearly moved past proclamations that 'privacy is dead.' Rather, privacy for operators of big data systems–much like environmental pollution for smokestack production facilities—has become a rapidly rising restraint that firms need to address in a sustainable way.

Emerging best practices in privacy design and engineering also align with a BDMS model. For organizations, the objective of Privacy by Design is to gain "a *sustainable* competitive advantage" by practicing 7 Foundational Principles.[14] The first Privacy By Design principle, "Proactive not reactive; Preventative not Remedial," holds the same proactive pollution prevention focus as EMS. Privacy By Design also calls for companies to respect privacy by making privacy protection an integral part of the way they do business. Similarly, The Privacy Engineer's Manifesto observes:

> *"Too often the necessary controls and measures to protect personal information required by a process, application, or system are either ignored or bolted on at the 11th hour of development."*[15]

The privacy engineering of a service or product that is using personal information or risking to reveal identity is part of the engineering of the service or product. The book comprehensively outlines an integral way forward defining Privacy Engineering "as using engineering principles and processes to build controls and measures into processes, systems, components, and products that enable the authorized, fair, and legitimate processing of personal information."[16]

Finally, organizations that adopt an EMS-like model for their big data projects will be inherently optimized for agility. In this time of rapid change, management systems of agility have a higher fitness than management systems striving only for efficiency. Environmental management systems, Agile development, Dev Ops and Open Source Software are all part of a wider management agility revolution well

[14] See PBD 7 Foundation Principles, https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/

[15] The Privacy Engineer's Manifesto, Kindle loc 1172 of 8181

[16] Id at Kindle 1187 of 8181

underway in organizational management theory. In his book XLR8, leading organizational change author John Kotter talks about the need for organizations to develop a dual operating system where a hierarchy acts as a superstructure for collaborative, self-forming teams to pursue big opportunities.[17] Jim Whitehurst, the CEO of the leading open source software company Red Hat, writes in his book The Open Organization, "[c]entral planning takes too long and consumes too many resources."[18] General Stanley McCrystal in his book Team of Teams explains how the hierarchical organization perfected last century for efficiency in the industrial revolution needs to give way in this century to a team of teams optimized for agility in the rapidly changing opening decades of this century's information revolution.[19] In the face of this overwhelming trend toward collaborative and agile management models, a BDMS model seems to fit better the way the business world works today.

IV. Conclusion

Like the need for environmental protection from fossil fuels at scale, there is a need to protect the ecology of data at scale. Organizations can better inform their decision making by thinking of big data management systems in environmental terms. The data protection regulations and management models we chose today need to be aligned with emerging, collaborative project management and software development methodologies such as Agile and DevOps which will develop and continuously improve upon big data analytics use cases. The emerging field of big data management should learn from the nearly fifty years of environmental management and move directly to embrace a front-end, integrated EMS-like approach. This will allow organizations to facilitate big data's benefits, mitigate its risks, and so support the value of their data-driven initiatives and contribute to the long-term sustainability of the big data economy.

[17] John Kotter, XLR8

[18] Jim Whitehurst. The Open Organization, pg 1.

[19] See General Stanley McCrystal et al. Team of Teams.