Washington and Lee Law Review Online

Volume 72 | Issue 3 Article 4

3-31-2016

Big Data Sustainability: An Environmental Management Systems Analogy

Dennis D. Hirsch
Ohio State College of Law

Jonathan H. King
Washington University

Big Data Sustainability: An Environmental Management Systems Analogy

Dennis D. Hirsch*
Jonathan H. King**

Abstract

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 management approach should businesses employ sustainably to achieve the tremendous benefits of big data analytics, while minimizing the potential negative externalities?

This Paper argues that leaders can learn from environmental management practices developed to manage the negative externalities of the industrial revolution. First, it shows that, along with its many benefits, big data can create 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 sustainably to develop their personal data assets. Second, this Paper chronicles environmental management's

* Professor of Law and Director of the Program on Data Law, Ethics, and Policy, Ohio State Moritz College of Law; Professor of Law, Capital University Law School.

** Visiting Scholar, Washington University and Head of Cloud Strategy, Ericsson. 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, Jared Wray, and the participants in the Future of Privacy Forum's *Beyond IRB's: Designing Ethical Review Processes for Big Data Research* event.

historical progression from a back-end, siloed approach to a more proactive and collaborative "environmental management system" method. An approach modeled after environmental management systems—a Big Data Management System approach—offers an effective model for managing data analytics operations to prevent negative externalities.

Finally, this Paper shows that a Big Data Management System approach aligns with: (A) Agile software development and DevOps practices that companies use to develop and maintain big data applications, (B) best practices in Privacy by Design and Privacy Engineering, and (C) emerging trends in organizational management theory. At this critical, formative moment when organizations begin to leverage personal data to revolutionary ends, we can readily learn from environmental management systems to embrace sustainable big data management from the outset.

Table of Contents

"We should work on our process, not the outcome of our processes." W. Edwards Deming[1]

Introduction

It is commonly proclaimed that "big data is the new oil."[2] This is true in the sense that data, like oil, constitutes a critical,

1. M. SCOTT CAMPBELL, PCISTM—ADVANCED PROJECT MANAGEMENT 70 (2016).

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,[3] identity pollution,[4] and harmful discrimination.[5] In both contexts, uses at scale produce not only tremendous societal benefits, but also meaningful, unintended 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

---

2. *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/ (last visited Feb. 7, 2016) ("Just like oil was a natural resource powering the last industrial revolution, data is going to be the natural resource for this industrial revolution.") (on file with the Washington and Lee Law Review).

3. *See* Dennis D. Hirsch, *The Glass House Effect: Big Data, the New Oil, and the Power of Analogy*, 66 ME. L. REV. 373, 375 (2014) (examining "the underside of the 'Big Data is the new oil' comparison").

4. *See* Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 43–45 (2013) (explaining the "identity paradox": "Big data seeks to identify, but it also threatens identity"); Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 422–26 (2014) (exploring how big data can compromise identity); Neil M. Richards & Jonathan H. King, *Big Data and the Future for Privacy*, *in* HANDBOOK OF RESEARCH ON DIGITAL TRANSFORMATIONS (forthcoming 2016) (manuscript at 8–10), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2512069 (describing how big data affects individuals' identities).

5. *See* Dennis D. Hirsch, *That's Unfair! Or is it? Big Data, Discrimination and the FTC's Unfairness Authority*, 103 KY. L.J. 345, 346 (2015) (stating that big data predictions "can result in unfair discrimination when the disfavored attributes further correlate to a particular race, religion, gender or other protected class so that the model ends up denying important life opportunities to people in these vulnerable groups").

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 the foregoing analogy, this Article proposes that Environmental Management Systems (EMS) provide a good reference model for organizations to consider for managing their expanding big data operations.

This Article makes this case in three parts. Part I shows that, along with its many benefits, big data can create 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 sustainably to develop their personal data assets. Part II chronicles environmental management's historical progression from a back-end, siloed approach to a more proactive and collaborative "environmental management system" method. Part II also responds to the idea that Consumer Subject Review Boards (CSRBs)—inspired by Internal Review Boards (IRBs)—constitute a useful model for big data management.[6] It explains that CSRB's 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 approach—would be integrated instead of compartmentalized; preventative rather than reactive. It offers a more effective model for managing data analytics operations to prevent negative externalities. Finally, Part III shows that a Big Data Management System approach aligns with: (A) Agile software development and DevOps practices that companies use to develop and maintain big data applications, (B) best practices in Privacy by Design and engineering, and (C) emerging trends in organizational management theory. These connections suggest that a systems oriented approach is a more 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

6. *See generally* Ryan Calo, *Consumer Subject Review Boards: A Thought Experiment*, 66 STAN. L. REV. ONLINE 97 (2013).

environmental management has traversed, from the time that companies paid little attention to their environmental impacts to the present day when many advanced companies compete to make their products and operations more environmentally and socially responsible. At this critical, formative moment when organizations begin to leverage personal data to revolutionary ends, we can readily learn from environmental management systems to embrace *sustainable* big data management from the outset.

I. Privacy Injuries Are to Big Data, As Pollution Is to Industrial Production

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.[7] This noble idea soon ran into problems. Parents of the schoolchildren worried that the 400 fields of data inBloom was collecting about students, including information on family violence, student disabilities, and other topics that might cast their children in a negative light, might attach to their children as they moved through life and constrain their educational and employment opportunities.[8] The parents grew concerned about who else would gain access to this data, either when inBloom intentionally shared data with others or if inBloom suffered from a data security breach. Parents began to protest

7. *See* Elizabeth Dwoskin & Lisa Fleisher, *Parental Opposition Fells inBloom Education-Software Firm: Privacy Concerns Over Use of Student Data Lead Company to Close*, WALL ST. J. (Apr. 21, 2014, 10:14 PM), http://www.wsj.com/articles/SB10001424052702304049904579516111954826916 (last visited Feb. 7, 2016) (describing how parental opposition caused schools to withhold student data and so forced inBloom to close) (on file with the Washington and Lee Law Review).

8. Natasha Singer, *InBloom Student Data Repository To Close*, N.Y. TIMES, April 22, 2014, at B2.

inBloom's collection and use of data about their children. School districts, and then entire states, refused to share student information with inBloom.[9] 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. The authors wrote this Article for the Future of Privacy Forum's "Beyond IRBs: Ethical Review Processes for Big Data Research" symposium.[10] The call for papers for this symposium focused on whether CSRBs, a management model based on IRBs and first proposed by Professor Ryan Calo in 2013,[11] could provide a useful model for data analytics governance. This Paper frames the question more broadly: What management approach should businesses employ sustainably to achieve the tremendous benefits of big data analytics, while minimizing the potential negative externalities? 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 compliance and found ways to turn environmental performance into competitive advantage by making more environmentally friendly products,[12] building trust in their brand,[13] and reducing regulatory costs.[14] While the

---

9. *Id.*

10. *See Beyond IRBs: Ethical Review Processes for Big Data Research*, FUTURE OF PRIVACY FORUM, https://bigdata.fpf.org/ (last visited Feb. 7, 2016) (describing this FPF event) (on file with the Washington and Lee Law Review).

11. *See generally* Calo, *supra* note 6.

12. *See* Dennis D. Hirsch, *Green Business and the Importance of Reflexive Law: What Michael Porter Didn't Say*, 62 ADMIN. L. REV. 1063, 1073–74 (2010) (providing examples).

13. *See id.* at 1079–80 (same).

14. *See id.* at 1081–82 (same).

transition from environmental compliance to using environmentally-friendly products and services for competitive advantage is far from complete, it has already generated valuable strategies for maximizing production benefits and minimizing negative externalities. The point is this: leaders who are tasked today with implementing big data projects and technologies can learn something from the development of environmentally-conscious strategies and management practices.

The environmental analogy also helps managers to perceive their situation more quickly and completely. 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 whether big data concerns are just a more complicated version of the over-exaggerated "Y2K" alarms. The environmental analogy can enable technical and nontechnical, legal and non-legal, commercial and governmental leaders to come to grips with how best to realize the benefits of big data analytics while managing the potential negative externalities. We can gain useful insights and add a measure of predictability by learning from the environmental policy, regulation, and management precedents of 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.[15]

This type of back-end environmental management strengthened compliance by the book but hurt production and stifled innovation in environmental protection itself. Environmental managers convinced product groups to fix mistakes and oversights in order to meet legal requirements after the fact. 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. Because most of the product and process design planning had occurred long before the environmental manager got involved, companies missed opportunities for upstream solutions—such as choices about product or process design or raw materials—that could have prevented the pollution from being created in the first place and addressed the issue at a 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 U.S. 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, rather than by fixing defects at the end of the line. The American automakers, 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 (TPS) that produced dramatic quality improvements while reducing costs and improving customer satisfaction. Eventually, TPS came to influence Total Quality Management (TQM) and, more recently, Lean Manufacturing.

---

15. *See generally* Dennis Hirsch, *How To Improve Privacy Protection by Adapting and Using Environmental Management Tools*, 5 PRIVACY OFFICERS ADVISOR 1 (2005).

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 TQM principles to the environmental arena was the Environmental Management System.[16]

Environmental Management Systems 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 EMSs demonstrate their ability to promote pollution prevention, enhance compliance, and reduce compliance costs.[17]

III. Big Data Management Systems

Viewed from the perspective of environmental management, a Consumer Subject Review Board management approach, and

16. Environmental management systems thus have a direct connection to Total Quality Management and, hence, to the Toyota Production System and Demming's theories.

17. *See generally* CHRISTOPHER SHELDON & MARK TOXON, ENVIRONMENTAL MANAGEMENT SYSTEMS: A STEP-BY-STEP GUIDE TO IMPLEMENTATION AND MAINTENANCE (2006) (providing an in-depth description of the benefits of EMSs).

the IRBs 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 IRBs themselves have engendered in the university context.

Data analytics management should take a lesson from environmental management. An EMS-like model—a Big Data Management System approach—would have data scientists, programmers and privacy professionals collaborating together so as 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 Big Data Management System 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 on any given big data project. This manager would ensure that product design, engineering, and operations teams see not only the benefits of their algorithmic creations, but also the privacy and discrimination issues that they may pose. This would reduce the need for late-stage evaluation of the product because societal implications—both beneficial and potentially harmful—would be considered throughout the process. Just as EMSs help prevent pollution, Big Data Management Systems should help prevent privacy and discriminatory impacts. Just as pollution prevention is less costly than end-of-pipe pollution controls, 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 applications and big data systems. Companies have increasingly moved from top-down, compartmentalized models such as "waterfall" to adopt Agile project management and DevOps software development methods that embrace an emergent and

collaborative approach.[18] 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.[19] A "minimum viable product" is conceived, launched, and then rapidly iterated upon by teams of people to improve as they operate.[20] 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- and fairness-related improvements as they arise.

A management system 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,[21] or simply to meet. In the Phoenix Project, a leading book on DevOps, 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

---

18. *See Principles Behind the Agile Manifesto*, MANIFESTO FOR AGILE SOFTWARE DEVELOPMENT, http://www.agilemanifesto.org/principles.html (last visited Feb. 8, 2016) (listing the twelve principles of the Agile Manifesto) (on file with the Washington and Lee Law Review).

19. *See* Dan Woods, *Why Lean and Agile Go Together*, FORBES (Jan. 12, 2010, 6:10 AM), http://www.forbes.com/2010/01/11/software-lean-manufacturing-technology-cio-network-agile.html (last visited Feb. 7, 2016) ("Agile development is an evolutionary conversation in which incremental steps of two to four weeks lead to feedback that allows requirements to be tested and adjusted.") (on file with the Washington and Lee Law Review); *see also* MARY POPPENDIECK & TOM POPPENDIECK, THE LEAN MINDSET: ASK THE RIGHT QUESTIONS 48–50 (2014) (outlining agile software development).

20. *See Methodology*, THE LEAN STARTUP, http://theleanstartup.com/principles (last visited Feb. 8, 2016) (describing the origination and meaning of the term "MVP") (on file with the Washington and Lee Law Review).

21. *See* William Boston, *Volkswagen Shares Dive on New Emissions Woes*, WALL ST. J. (Nov. 4, 2015, 3:23 AM), http://www.wsj.com/articles/volkswagens-shares-take-tumble-after-epas-fresh-allegations-1446559388 (last visited Feb. 7, 2016) (illustrating the consequences of evading environmental regulations) (on file with the Washington and Lee Law Review).

failure to address restraints causes unplanned work, which breaks sustainable operations: "Your job as VP of IT Operations is to ensure the fast, predictable, and uninterrupted flow of planned work that delivers value to the business while minimizing the impact and disruption of unplanned work, so you can provide stable, predictable, and secure IT service."[22] 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 smart and sustainable way.

Emerging best practices in privacy by design and engineering also align with a Big Data Management System approach. For organizations, the objective of Privacy by Design is to gain "a *sustainable* competitive advantage" by practicing seven Foundational Principles.[23] 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 make 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."[24] The privacy engineering of a service or product that is using personal data or risking revealing identity is part of the engineering of the service or product. The Manifesto defines 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

22. GENE KIM, KEVIN BEHR & GEORGE SPAFFORD, THE PHOENIX PROJECT: A NOVEL ABOUT IT, DEVOPS, AND HELPING YOUR BUSINESS WIN 91 (2014).

23. *See* ANN CAVOUKIAN, PRIVACY BY DESIGN: THE 7 FOUNDATIONAL PRINCIPLES, INFO. & PRIVACY COMM'R OF ONT. (2011), https://www.ipc.on.ca/images/Resources/7foundationalprinciples.pdf (listing these seven principles).

24. MICHELLE FINNERAN DENNEDY, JONATHAN FOX & THOMAS FINNERAN, THE PRIVACY ENGINEER'S MANIFESTO: GETTING FROM POLICY TO CODE TO QA TO VALUE 66 (2014).

information."[25] Both Privacy by Design and Privacy Engineering seek to get employees from various departments to make privacy a core part of their jobs, thereby enabling the organization to build privacy into its products and processes from the beginning instead of addressing it later.

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, agile management systems have a higher fitness than those that that seek principally to streamline management processes and make them more efficient. Environmental management systems, Agile development, DevOps, and Open Source Software are all part of a wider agility revolution well underway in organizational management theory. In his book *Accelerate*, 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.[26] Jim Whitehurst, the CEO of the leading open source software company Red Hat, writes in his book *The Open Organization*, "Central planning takes too long and consumes too many resources."[27] General Stanley McChrystal 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.[28] In the face of this overwhelming trend toward collaborative and agile management, a Big Data Management System model fits better with the way the business world works today.

---

25. *Id.* at 29.

26. *See* JOHN KOTTER, ACCELERATE: BUILDING STRATEGIC AGILITY FOR A FASTER-MOVING WORLD 19–39 (2014) (delineating the structure and pros of dual operating systems).

27. JIM WHITEHURST, THE OPEN ORGANIZATION: IGNITING PASSION AND PERFORMANCE 1 (2015).

28. *See generally* GENERAL STANLEY MCCHRYSTAL ET AL., TEAM OF TEAMS: NEW RULES OF ENGAGEMENT FOR A COMPLEX WORLD (2015) (setting forth this argument).

IV. Conclusion

Like the need for environmental protection from fossil fuels at scale, there is a need to protect the ecology of personal 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 choose 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, support the value of their data-driven initiatives and contribute to the long-term sustainability of the big data economy.