Big Data Sustainability

An Environmental Management Systems Analogy

Dennis D. Hirsch[[1]](#footnote-1)\*

Jonathan H. King[[2]](#footnote-2)\*\*

**Introduction**

It is commonly proclaimed that “big data is the new oil.”[[3]](#footnote-3) 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 involve data spills, privacy violations, and identity pollution[[4]](#footnote-4). In both contexts, uses at scale produce not only tremendous societal benefits, but also meaningful societal externalities that can produce unintended consequences and run afoul of regulators.[[5]](#footnote-5)

Comparing big data to oil is but one of many timely analogies to be made between the environmental externalities of the industrial revolution and the emerging data externalities of the information revolution. At this formative moment of mass big data adoption, we can learn from environmental management practices developed to manage 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 adapting an Environmental Management System or “EMS” is a good reference model for organizations to consider for managing their big data developments.

We support our proposal by first further examining the utility of the analogy between the externalities of the information revolution and the industrial revolution. We then summarize the evolution of traditional environmental management from a siloed, command and control structure to a more collaborative, environmental management system approach. Finally, we argue why an environmental management system provides a good reference case for big data management. An EMS-like approach, which we call a Big Data Sustainability System or “BDSS,” is a natural fit for big data management because it naturally 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. At this critical moment when organizations want to leverage big data to revolutionary ends, we can learn from environmental management systems and sooner achieve big data sustainability.

1. **Big Data Environmental Analogy**

We live in a time of accelerating but precedented change. Accelerating in that the information revolution, powered by the digital revolution of Moore’s law, has our society and economy hurtling forward at accelerating rates. Precedented, in that the digital revolution comes in the wake of the industrial revolution which in the span of two centuries has utterly changed, and continues to change, our society and planet as profoundly as the information revolution will. Change in that the confluence of the industrial revolution and the pace of the information revolution see those of us with children wondering where it is all heading for their children.

In times of change, analogies are powerful. Professors Gavetti and Rivkin in their Harvard Business Review Article, *How Strategists Really Think: Tapping the Power of Analogy*, explain how analogies help us understand contemporary problems and enable us to develop solutions to them.[[6]](#footnote-6) Our analogy simply said at a high level is that just as the industrial revolution gave rise to environmental protection, the information revolution is giving rise to data protection. We believe this analogy is helpful to both understand the significance of big data and to develop solutions to manage potential negative externalities.

Just expanding upon the ‘big data is the new oil’ analogy to examine the externalities surrounding fossil fuels, we can gain context for today’s high stakes surrounding big data. Harnessing oil and other fossil fuels stored in the ground unleashed industrial innovation and productivity enabling new feats of human ingenuity to build and sustain the industrial economy we depend on today. The use and development of fossil fuels required complex systems of extraction, distribution and production involving trillions of dollars of investment.[[7]](#footnote-7) The uses of fossil fuels are so revolutionary and occur at such a scale as to impact the environment in ways that protecting the environment itself became a constraint of sustained fossil fuel use for industrial development. Hence the sustainable success of the industrial revolution has come to involve not just industrial innovations to extract, refine and use fossil fuels but environmental protection innovations as well. Globally, new environmental policies, regulations, management models, inventions, and societal norms emerged throughout the 20th century to begin to sustainably reconcile industrial development with environmental protection.

Now compare this to where we sit in the information revolution of the 21st century. We have entered a ‘critical window’ where society similarly debates how to sustainably reconcile the mass adoption of big data technologies with data protection principles.[[8]](#footnote-8) Governments and organizations across the world wrestle with how to extract valuable insights from diverse data sets without invading privacy, compromising identity or otherwise undermining sustainable big data use. Much like managing fossil fuel extraction and consumption, government regulators face impactful trade offs for their economies and societies in how they chose to protect and regulate big data use in their jurisdictions. Much like with the development and use of fossil fuels, organizations have already made and prepare to make billions of dollars worth of investment decisions to unleash the power of big data analytics. Globally, privacy policies, regulations, inventions, societal norms and management models to sustainably reconcile big data use with data protection are unclear and rapidly changing.

Since the rate of change is accelerating in the information revolution, data protection decisions are playing out more rapidly. Episodic tumult such as Edward Snowden’s disclosures and the European Court of Justice’s decision to invalidate the Safe Harbor complicate data protection policy making and big data investment decisions.[[9]](#footnote-9) 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 an environmental big data 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. This is especially true of the evolution of environmental management systems.

1. **Environmental Management System**

Prior to the emergence of EMS’s, environmental managers operated in a reactive, case-by-case and compartmentalized traditional approach to enforce environmental compliance. The design and production departments of the business decided what they wanted to make and how they would produce it. Then, towards the end of the planning process, the environmental manager would tell the business what it needed to do to comply with environmental laws. All too often, this meant telling the business what environmental regulations would not let the business do, and then sending the business teams back to rework their plans. This type of back-end environmental management strengthened compliance by the book but stifled innovation in environmental compliance itself. Environmental managers had to retroactively get product groups to fix mistakes and oversights to meet legal requirements. This took more time and, when it required that plans be changed, imposed major delays. Institutions viewed environmental managers as internal cops and environmental management as a necessary evil.[[10]](#footnote-10)

While environmental management continued on this course, the broader management of industrial mass production began to change. Catalyzed by the work of statistician W. Edwards Deming, new production approaches emerged in the auto industry. Deming examined traditional management methods for ensuring quality where teams fixed defects at the end of production lines. Deming advocated improving quality by thinking of manufacturing as a system to continuously improve, not as bits and pieces to inspect and fix after assembly. 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 to be known as Total Quality Management and more recently as Lean Manufacturing.

Forward-thinking companies started to consider whether Total Quality Management could be adapted for 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. This application of Total Quality Management to the environmental arena lead to the development of Environmental Management Systems.

Environmental Management Systems differ significantly from traditional environmental management. Instead of being compartmentalized, 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 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 being reactive to control pollution, collaborative teams innovate to prevent pollution. Frequently, these front-end, pollution prevention solutions end up saving organizations money as compared to end-of-line controls. Accordingly, studies of EMS’s demonstrate their ability to promote pollution prevention, enhance compliance and reduce compliance costs.[[11]](#footnote-11)

1. **Big Data Sustainability Systems**

We believe that big data management can achieve analogous benefits by adopting an environmental management systems approach. An EMS-like model which we will call a Big Data Sustainability System or “BDSS” 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 BDSS management structure would have the person responsible for mitigating privacy and discriminatory impacts present at the front end of the process as part of an 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 issues as they design and implement algorithms and applications. This would eliminate 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 BDSS’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 than a cumbersome review process at the back end.

An approach grounded in the EMS model fits naturally with the way that companies increasingly test, develop and operate their big data systems and applications. Companies have moved from command-and-control models such as ‘water-fall’ to increasingly adopt Agile project management and DevOps software development methods that embrace an emergent and collaborative approach.[[12]](#footnote-12) Originating from the same Lean Manufacturing roots as EMS, Agile and DevOps seek to make continuous improvements throughout the process, not at the end of it. [[13]](#footnote-13) A ‘minimum viable product’ is conceived, launched and then rapidly iterated upon by teams of people to improve as they operate.[[14]](#footnote-14) 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 BDSS model is naturally aligned with Agile and DevOps mindsets. Privacy can be seen as an engineering restraint to continuously improve upon, not deny, avoid or evade.[[15]](#footnote-15) 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.”[[16]](#footnote-16)*

In a post Snowden era with no more Safe Harbor, we have clearly moved past proclamations that ‘privacy is dead.’ Rather, privacy is a rapidly rising restraint for operators of IT systems, especially including big data systems, to sustainably address.

Emerging best practices in privacy design and engineering also align with a BDSS model. For organizations, the objective of Privacy by Design is to gain “a **sustainable** competitive advantage” by practicing 7 Foundational Principles.[[17]](#footnote-17) 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.”[[18]](#footnote-18)*

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.”[[19]](#footnote-19)

Finally, organizations who 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 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.[[20]](#footnote-20) 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.”[[21]](#footnote-21) 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.[[22]](#footnote-22) In the face of this overwhelming trend toward collaborative and agile management models, an IRB approach would risk being a generation behind the way the world works today. This at a time when we need a big data management approach that is thinking generations ahead.

1. **Conclusion**

Like the need for environmental protection from fossil fuels at scale, there is a need for ecology of data at scale, its called data protection. 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. 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. The emerging field of big data management should learn from nearly fifty years of environmental management and move directly to embrace a front-end, integrated EMS-like approach. Big data sustainability is crucial at this tipping point not only for our information based society, but the sustainability of our industrial society as well. Big data and data protection can more predictably and sustainably mature by learning from environmental management. In return maybe sustainable big data analytics can revolutionize sustainable industrial development in time to help save the planet.

1. \* Geraldine W. Howell Professor of Law, Capital University Law School. [↑](#footnote-ref-1)
2. \*\* 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. [↑](#footnote-ref-2)
3. *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/. [↑](#footnote-ref-3)
4. See Dennis D. Hirsch, *The Glass House Effect: Big Data, The New Oil, And The Power OF Analogy*, 66 Maine Law Review, 374 (2014). [↑](#footnote-ref-4)
5. *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). [↑](#footnote-ref-5)
6. *See* Giovanni Gavetti & Jan W. Rivkin *How Strategists Really Think: Tapping the Power of Analogy, Harv. Bus. Rev., Apr. 1, 2005 at 1. See also D. Hirsch, The Glass House Effect* [↑](#footnote-ref-6)
7. *See* Tom Randall, *Bankers See $1 Trillion of Zombie Investments Stranded in the Oil Fields,* Bloomberg Business (Dec. 17, 2014), http://www.bloomberg.com/news/articles/2014-12-18/bankers-see-1-trillion-of-investments-stranded-in-the-oil-fields. [↑](#footnote-ref-7)
8. *See* [European Data Protection Supervisor, Towards A New Digital Ethics, Data, Dignity & Technology, Opinion 4/2015 (Sep. 11, 2015) available at <https://secure.edps.europa.eu/EDPSWEB/webdav/site/mySite/shared/Documents/Consultation/Opinions/2015/15-09-11_Data_Ethics_EN.pdf>; Citing Richards & King, supra note x. [↑](#footnote-ref-8)
9. See Mark Scott, Data Transfer Pact Between U.S. and Europe Is Ruled Invalid, N.Y. Times (Oct. 6, 2015) [↑](#footnote-ref-9)
10. See Dennis Hirsch, How To Improve Privacy Protection by Adpating and Using Environmental Management Tools, Privacy Officers Advisor, Aug. 2005, Vol 5, No 11. [↑](#footnote-ref-10)
11. See Christopher Sheldon & Mark Toxon, Environmental Management Systems (2006). [↑](#footnote-ref-11)
12. See Agile Manifesto – 12 principles and commentary. [↑](#footnote-ref-12)
13. 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 [↑](#footnote-ref-13)
14. See Lean Startup, origination of term “MVP”. [↑](#footnote-ref-14)
15. See William Boston, Volkswagon Shares Dive on Emission Woes, The Wall Street Jounral (Nov. 4, 2015). [↑](#footnote-ref-15)
16. The Phoenix Project, Kindle 1340 of 5829 [↑](#footnote-ref-16)
17. See PBD 7 Foundation Principles, https://www.privacybydesign.ca/index.php/about-pbd/7-foundational-principles/ [↑](#footnote-ref-17)
18. The Privacy Engineer’s Manifesto, Kindle loc 1172 of 8181 [↑](#footnote-ref-18)
19. Id at Kindle 1187 of 8181 [↑](#footnote-ref-19)
20. John Kotter, XLR8 [↑](#footnote-ref-20)
21. Jim Whitehurst. The Open Organization, pg 1. [↑](#footnote-ref-21)
22. See General Stanley McCrystal et al. Team of Teams. [↑](#footnote-ref-22)