2017

Technological Innovation, Data Analytics, and Environmental Enforcement

David L. Markell
Florida State University College of Law

Robert L. Glicksman
George Washington University Law School

Claire Monteleoni
George Washington University

Technological Innovation, Data Analytics, and Environmental Enforcement

Robert L. Glicksman, David L. Markell,** and Claire Monteleoni****

Technical innovation is ubiquitous in contemporary society and contributes to its extraordinarily dynamic character. Sometimes these innovations have significant effects on the environment or on human health. They may also stimulate efforts to develop second-order technologies to ameliorate those effects. The development of the automobile and its impact on life in the United States and throughout the world is an example. The story of modern environmental regulation more generally includes chapters filled with examples of similar efforts to respond to an enormous array of technological advances.

This Article uses a different lens to consider the role of technological innovation. In particular, it considers how technological advances have the potential to shape governance efforts in the compliance realm. The Article demonstrates that such technological advances—especially new and improved monitoring capacity, advances in information dissemination through e-reporting and other techniques, and improved capacity to analyze information—have significant potential to transform governance efforts to promote compliance. Such transformation is likely to affect not only the "how" of compliance promotion, but also the "who"—who is involved in promoting

DOI: https://dx.doi.org/10.15779/Z38GX44T46

* J.B. & Maurice C. Shapiro Professor of Environmental Law, The George Washington University Law School. This Article is an effort at interdisciplinary collaboration. Two of us are law professors with extensive experience in the environmental compliance arena. Our third co-author is an expert in the world of information systems, including data analytics and the use of "big data." This Article seeks to bring together our respective areas of expertise. We are grateful to Professors Greg Mandel and J.B. Ruhl for their thoughtful comments on drafts of the Article. Katie Miller, Assistant Director for ILL & Document Delivery, Florida State University College of Law Research Center, provided very helpful research assistance.

** Steven M. Goldstein Professor of Law and Associate Dean for Research, Florida State University College of Law.

*** Assistant Professor, Department of Computer Science, The George Washington University.

compliance. Technological innovation is likely to contribute to new thinking about the roles key actors can and should play in promoting compliance with legal norms. The Article discusses some of the potential benefits of these types of technological innovation in the context of the Environmental Protection Agency's ongoing efforts to improve its compliance efforts by taking advantage of emerging technologies. We also identify some of the pitfalls or challenges that agencies such as the Environmental Protection Agency need to be aware of in opening this emerging bundle of new tools and making use of them to address real-world environmental needs.

INTRODUCTION

The development of increasingly sophisticated information technology presents opportunities and challenges for the operation of the regulatory state. This Article addresses those dynamics in the context of environmental enforcement, and in particular, assesses the promise and potential pitfalls of the efforts of the federal Environmental Protection Agency (EPA) to harness new information technology and advanced data analytics to transform its enforcement programs through greater reliance on advanced monitoring and reporting technologies.

Technological innovation has played a critical role in how our country has developed. Inevitably, such innovation has similarly played a central role in the development of environmental law and policy. Take for instance, the internal combustion engine, which enabled the automobile to transform society economically and socially.[1] This transformative technological innovation, however, generates air pollution, which society determined required regulatory attention under environmental laws.[2] In the mid-1960s, Congress first adopted provisions,[3] whose current version is reflected in Title II of the Clean Air Act (CAA),[4] authorizing federal regulations to address pollution from mobile sources such as automobiles. It required the then-Secretary of Health, Education, and Welfare to adopt emission control standards taking into account technological feasibility and cost.[5] In doing so, the legislature assumed that the automobile industry would use catalytic converters (a second category of technologies) to limit the harmful effects of the internal combustion engine (the initial technological innovation).[6]

As the development of motor vehicle emissions standards demonstrates, Congress, in selecting approaches from its regulatory tool box,[7] has frequently required that regulated entities comply with standards that specify levels of performance that are achievable using the best available technology within their industry.[8] Technological development is not static, however, and neither is regulation that responds to or relies on it. Regulation of air pollution, for

1. "Perhaps no invention affected American everyday life in the 20th century more than the automobile." *Pre-Columbian to the New Millennium: The Age of the Automobile*, USHISTORY.ORG, http://www.ushistory.org/us/46a.asp (last visited Jan. 9, 2007). Among other things, the invention of the automobile created new industries and jobs in fields such as rubber production and road construction, and immeasurably increased mobility. *Id.*; *see also* Anya C. Musto, Comment, *California as a Model for Federal Regulation of Automobile Emissions Pollution: Replacing Title II of the Clean Air Act of 1990*, 5 DICK. J. ENVTL. L. & POL'Y 151, 152 (1996) ("The prominence of the automobile in American society defines the problem of air pollution in both personal and political terms. The American automobile sustains the economy, dictates land use patterns, and provides freedom for the population.").

2. *See* Sanya Carley et al., *Innovation in the Auto Industry: The Role of the U.S. Environmental Protection Agency*, 21 DUKE ENVTL. L. & POL'Y F. 367, 370 (2011) (discussing efforts to "retain the internal combustion engine but power it with a petroleum substitute that can be produced in the United States and accomplish a more acceptable profile of environmental effects").

3. The Motor Vehicle Air Pollution Control Act, Pub. L. No. 89-272, 79 Stat. 992 (1965); Air Quality Act of 1967, Pub. L. No. 91-137, 83 Stat. 283 (1967).

4. 42 U.S.C. §§ 7521–7554 (2012).

5. Pub. L. No. 89-272, § 101(8), 79 Stat. 992 (1965).

6. *See* Thomas O. McGarity, *MTBE: A Precautionary Tale*, 28 HARV. ENVTL. L. REV. 281, 294 (2004) ("In prescribing tailpipe emissions standards, Congress in 1970 assumed that the automobile manufacturing industry would meet those standards by installing catalytic converters in the exhaust stream."); *cf.* Andrew P. Morriss, *The Next Generation of Mobile Source Regulation*, 17 N.Y.U. ENVTL. L.J. 325, 346 (2008) (noting that the 1970 CAA "authorized the EPA to order refiners to alter gasoline formulations to protect the catalytic converters").

7. *See* U.S. CONGRESS, OFFICE OF TECH. ASSESSMENT, ENVIRONMENTAL POLICY TOOLS: A USER'S GUIDE, OTA-ENV-634 8–21 (1995), https://www.princeton.edu/~ota/disk1/1995/9517/9517.PDF.

8. *See, e.g.*, 33 U.S.C. §§ 1311(b), 1314(b) (2012) (Clean Water Act); 42 U.S.C. § 7411(a)(1), (b)(1)(B) (2012) (Clean Air Act).

example, has evolved over time as control technologies improved,[9] efforts to control one pollutant affected the feasibility of controlling others, increases in fuel efficiency affected emission control performance, and as we learned more about the risks involved.[10]

Some environmental laws do not limit regulators to the adoption of standards that track available control technologies. Rather, they are more ambitious, seeking to "force" the development of new and improved technologies to control adverse environmental spillovers by establishing regulatory requirements not yet achievable using currently available control techniques.[11] Some of these technology-forcing experiments, such as motor vehicle emission standards prompting the development of the catalytic converter,[12] have been successful. They have led to the discovery and

9. The CAA requires EPA to revise its emission standards for motor vehicles "from time to time" based on the same technological feasibility and cost considerations as apply to their initial adoption. 42 U.S.C. § 7521(a)(1) (2012).

10. *See* Arnold W. Reitze, Jr., *Mobile Source Air Pollution Control*, 6 ENVTL. LAW. 309, 325–27 (2000) (describing the relationships among these factors). One interesting feature of regulation of mobile source pollution has been Congress's decision to empower California to impose more stringent emission standards, and simultaneously other states to adopt either the federal or California approach. 42 U.S.C. §§ 7507, 7543(b) (2012). This decision has induced manufacturers who want to operate in the California market to build and market nationally cars capable of meeting that state's standards to avoid the inefficiencies of multiple production lines. *See* Laura Moore Smith, *Divided We Fall: The Shortcomings of the European Union's Proposal for Independent Member States to Regulate the Cultivation of Genetically Modified Organisms*, 33 U. PA. J. INT'L L. 841, 859 (2012) ("By allowing two different emission standards, manufacturers either have to build 'California standard cars' and 'federal standard cars,' or simply build cars for the more stringent California standards (thus making separate federal standards moot)."). Another contemporary example of the central role technological developments play is in the energy field, including the ongoing debate about hydraulic fracturing and its impact on the nation's energy mix and on the environment. *See generally* John M. Golden & Hannah J. Wiseman, *The Fracking Revolution: Shale Gas as a Case Study in Innovation Policy*, 64 EMORY L.J. 955 (2015); Richard J. Pierce, Jr., *Natural Gas: A Long Bridge to a Promising Destination*, 32 UTAH ENVTL. L. REV. 245 (2012); John Schwartz, *Another Inconvenient Truth: It's Hard to Agree How to Fight Climate Change*, N.Y. TIMES (July 11, 2016), http://www.nytimes.com/2016/07/12/science/climate-change-movement.html (noting that prominent environmental groups once praised natural gas as such a bridge, but that some such groups have shifted positions, now referring to natural gas as a "bridge to nowhere"). On another energy front, technological advances have lowered the cost of energy produced by solar and wind power, but these sources, too, leave an environmental footprint. *See generally* Robert L. Glicksman, *Solar Energy Development on the Federal Public Lands: Environmental Trade-Offs on the Road to A Lower-Carbon Future*, 3 SAN DIEGO J. CLIMATE & ENERGY L. 107 (2012); Melanie McCammon, *Environmental Perspectives on Siting Wind Farms: Is Greater Federal Control Warranted?*, 17 N.Y.U. ENVTL. L.J. 1243 (2009); Steven Ferrey, *Ring-Fencing the Power Envelope of History's Second Most Important Invention of All Time*, 40 WM. & MARY ENVTL. L. & POL'Y REV. 1, 11 (2015) (noting the "big change . . . ushered in through the technological and cost declines of wind and solar photovoltaic ('PV') distributed generation").

11. *See, e.g.*, Whitman v. Am. Trucking Ass'ns, 531 U.S. 457, 490 (2001) (Breyer, J., concurring in part and concurring in the judgment) (contending that the CAA's legislative history "shows that Congress intended the statute to be 'technology forcing'").

12. Dennis D. Hirsch, *Green Business and the Importance of Reflexive Law: What Michael Porter Didn't Say*, 62 ADMIN. L. REV. 1063, 1102 (2010).

implementation of new technologies that limit or avoid pollution.[13] Some have contended, however, that the extent to which technology-forcing mandates have generated technological innovation has been limited by factors such as the absence of adequate rewards for innovation.[14]

For some processes, the search for technological fixes has included enormous investments of time and money, but has not yet yielded hoped-for results. Disposal of spent fuel and other forms of high-level radioactive waste generated by operation of nuclear power plants is one example.[15] The search for technological fixes to greenhouse gas emissions that contribute to climate change from coal-fired power plants is another.[16] In each case, the inability to devise technological fixes to abate environmental and health concerns has had significant effects on our country's economy and energy mix.[17]

In addition to technologies that create environmental concerns and others that can help to ameliorate their adverse effects, a third category of technologies that are of foundational importance to environmental law are those that monitor environmental conditions. Some monitoring tools measure ambient environmental quality, while others track releases of pollution associated with regulated party operations. Monitoring both the current state of the environment and releases of pollutants assures (or at least promotes) compliance with environmental norms, which is central to achieving the goals of the environmental statutes.[18] For the same reason, a vigorous enforcement

---

13. *See* Thomas O. McGarity, *Radical Technology-Forcing in Environmental Regulation*, 27 LOY. L.A. L. REV. 943, 945–53 (1994) (citing as examples of successful statutory technology-forcing ventures the phase-out of the pesticide Mirex and the phase-out of lead in gasoline); Gaia J. Larsen, *Skewed Incentives: How Offshore Drilling Policies Fail to Induce Innovation to Reduce Social and Environmental Costs*, 31 STAN. ENVTL. L.J. 139, 167–68 (2012) (discussing the lead phase-out).

14. *See* Gregory N. Mandel, *Innovation Rewards: Towards Solving the Twin Market Failures of Public Goods*, 18 VAND. J. ENT. & TECH. L. 303, 304 (2016) ("Despite numerous and diverse efforts, one significant goal that has largely eluded environmental law is adequately promoting environmentally beneficial innovation. While there have been many attempts at technology-forcing and innovation-promoting legislation in jurisdictions around the world, success has been limited."); D. Bruce La Pierre, *Technology-Forcing and Federal Environmental Protection Statutes*, 62 IOWA L. REV. 771, 837–38 (1977) (discussing political pressure and other factors that reduce theoretical incentives for major innovation).

15. *See* 42 U.S.C. § 10131(a)(2)–(3) (2012) (finding that "a national problem has been created by the accumulation of (A) spent nuclear fuel from nuclear reactors; and (B) radioactive waste from (i) reprocessing of spent nuclear fuel; (ii) activities related to medical research, diagnosis, and treatment; and (iii) other sources" and that "[f]ederal efforts during the past 30 years to devise a permanent solution to the problems of civilian radioactive waste disposal have not been adequate").

16. *See, e.g.*, J.B. Ruhl, *The Political Economy of Climate Change Winners*, 97 MINN. L. REV. 206, 218 (2012) (discussing likely inability to develop effective greenhouse gas mitigation techniques and technologies, such as sequestration, within currently feasible planning horizons).

17. *See, e.g.*, Alex Funk & Benjamin K. Sovacool, *Wasted Opportunities: Resolving the Impasse in United States Nuclear Waste Policy*, 34 ENERGY L.J. 113, 114–15 (2013) (arguing that the failure of energy planners and electric utility operators to develop technologies for the long-term disposal of nuclear waste has been an "Achilles Heel" that has prevented a "nuclear renaissance" in the production of no-carbon energy).

18. *See, e.g.*, Daniel A. Farber, *Taking Slippage Seriously: Noncompliance and Creative Compliance in Environmental Law*, 23 HARV. ENVTL. L. REV. 297, 304 (1999) (contending that it is

presence is critical.[19] The completeness and accuracy of efforts to measure the extent of compliance have significant impacts on governance capacity and performance.[20] Such measurement efforts can be used to identify areas of environmental concern and plan environmental policy strategies. In the compliance arena, they can help with essential tasks, such as identifying regulated entities in violation of pollution control regulations or permits, and providing evidence in enforcement actions taken against those entities.[21]

In recent years, a revolution involving this third interface of technology and environmental law has occurred. New or better technologies have advanced the capacity of governmental and nongovernmental actors to identify, measure, share, analyze, report on, and respond to the effects of activities subject to environmental regulations. This development is but a small part of the explosion of information technology, which has increased society's capacity to generate and analyze data by orders of magnitude. By one account, "[t]he rapid evolution of cyberspace and the accompanying rise of Big Data[22] has clearly been one of the greatest technological revolutions in recorded history."[23] The changes in information analytics have dramatically affected public policy in diverse areas such as national security[24] and health care,[25] giving rise to a host of legal issues, including the need to fashion protections for personal privacy rights[26] and the potential for the use of data analytics to enable undesirable

"obvious that translating legal mandates into actual compliance is far from automatic" and that insufficient attention has been paid to slippage between enactment of and compliance with environmental laws).

19. Cynthia Giles, *Next Generation Compliance*, 30 ENVTL. F. 22, 22 (Sept./Oct. 2013) (arguing that "strong criminal and civil enforcement is—and will continue to be—an essential part of [EPA's] environmental protection work").

20. *See* David Markell, *An Overview of TSCA, Its History and Key Underlying Assumptions, and Its Place in Environmental Regulation*, 32 WASH. U. J.L. & POL'Y 333, 375 n.4 (2010) ("The questions of how we should measure progress and, related, the metrics we should use to gauge success, are important parts of this debate that remain unsettled.").

21. For a survey of the uses of monitoring technologies and modeling programs under the federal CAA, see ROBERT L. GLICKSMAN ET AL., ENVIRONMENTAL PROTECTION: LAW AND POLICY 479-481 (Wolters Kluwer 7th ed. 2015).

22. We address the meaning of the term big data below. *See infra* Part I.A.

23. Abraham R. Wagner & Paul Finkelman, *Security, Privacy, and Technology Development: The Impact on National Security*, 2 TEX. A&M L. REV. 597, 614 (2015).

24. *See, e.g.*, Stephen I. Vladeck, *Big Data Before and After Snowden*, 7 J. NAT'L SECURITY L. & POL'Y 333, 333 (2014).

25. *See, e.g.*, Janine S. Hiller, *Healthy Predictions? Questions for Data Analytics in Health Care*, 53 AM. BUS. L.J. 251, 251 (2016).

26. *See, e.g.*, Daniel J. Solove, *Privacy and Power: Computer Databases and Metaphors for Information Privacy*, 53 STAN. L. REV. 1393, 1393 (2001); Joseph Jerome, *Big Data: Catalyst for a Privacy Conversation*, 48 IND. L. REV. 213, 213–14 (2014); Kevin P. Brady, *Student Cell Phone Searches and Reasonable Suspicion in a Digital Age: The Future Implications of* Riley v. California, 309 ED. LAW REP. 1, 14 (2014) (reporting that "inBloom, a nonprofit data analytics corporation financially supported by the Melinda and Bill Gates Foundation and the Carnegie Corporation decided to shut down its operations amid controversy and nationwide protests by parents and privacy advocates that the confidentiality of student digital records could not be properly protected in a 'cloud–based' computing environment").

government initiatives.[27] In the environmental compliance sphere, these technological advances have the potential to transform the capacities of government officials, regulated parties, and interested citizens.[28] They also have the potential to transform relationships between and among these actors, and the roles each performs.[29]

We suggest, in other words, that the development of technologies that generate new streams of data and enable new and better analyses does not merely have the potential to increase compliance with environmental law and improve environmental conditions—an ambitious agenda in its own right. In addition, these technological advances have the potential to significantly empower all of the relevant stakeholders in the environmental policy-making and implementation process and thereby play a significant role in transforming the governance landscape. Information technology can increase the efficiency and effectiveness of EPA's compliance and enforcement programs by providing relevant low-cost information, enabling the agency to reduce the levels of staffing needed to collect and process information, and facilitating the agency's redirection of its resources based on an improved understanding of compliance needs. Optimal use of big data, however, will require EPA to hire experts in data analytics[30] and make significant investments in computer systems capable of collecting, transporting, storing, and analyzing the data.[31]

---

27. *See, e.g.*, Kelsey Finch & Omer Tene, *Welcome to the Metropticon: Protecting Privacy in a Hyperconnected Town*, 41 FORDHAM URB. L.J. 1581, 1603 (2014) (noting that New York City's "data-driven stop-and-frisk policy was accused by critics of actively targeting African American and Hispanic residents"). Government's ability to respond effectively to changes in information technology, like other examples of technological innovation, is likely to be affected by the so-called "pacing problem," which results from the development of technological innovation at a pace faster than the development of appropriate regulatory responses. *See, e.g.*, Wulf A. Kaal, *Dynamic Regulation for Innovation*, *in* PERSPECTIVES IN LAW, BUSINESS & INNOVATION (Mark Fenwick et al. eds., New York Springer, 2016) (forthcoming), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2831040.

28. For examples of the integration of big data into environmental law and policy initiatives, see Linda K. Breggin & Judith Amsalem, *Big Data and the Environment: A Survey of Initiatives and Observations Moving Forward*, 44 ENVTL. L. REP. 10,984 (2014).

29. *See, e.g.*, Giles, *supra* note 19, at 24 (predicting that "changes, driven by new technologies, will encourage more direct industry and community engagement, and reduce the need for government action"). For more detailed treatment of the roles of different actors, see David L. Markell & Robert L. Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming); David L. Markell & Robert L. Glicksman, *Dynamic Governance in Theory and Application, Part I*, 58 ARIZ. L. REV. 563, 618–29 (2016) [hereinafter *Dynamic Governance, Part I*]; David L. Markell & Robert L. Glicksman, *A Holistic Look at Agency Enforcement*, 93 N.C. L. REV. 1 (2014) [hereinafter, Markell & Glicksman, *A Holistic Look*].

30. *See, e.g.*, Jeff Chasney, Commentary, *Data Analytics Can Backfire Without Experts*, INFORMATIONWEEK (Jan. 10, 2012, 2:23 PM), http://www.informationweek.com/software/information-management/data-analytics-can-backfire-without-experts/d/d-id/1102175?; Bhavani Raskutti, *The Role of Domain Experts in Data Science*, TERADATA (Mar. 23, 2015), http://blogs.teradata.com/international/role-domain-experts-data-science/.

31. *See, e.g.*, Liane Colonna, *A Taxonomy and Classification of Data Mining*, 16 SMU SCI. & TECH. L. REV. 309, 329 (2013) (noting that "data mining differs from analytics because it 'uses more complex computer modeling, database analysis, and theoretical modeling which often requires a significant investment in software, computer hardware, and specialized data analysis resources'");

New data-gathering and processing capacity can also upgrade EPA-state relations by fostering information sharing and better coordinated enforcement activity between EPA and its state partners under the environmental cooperative-federalism statutes. In addition, it can improve regulated parties' capacity to identify, diagnose, and address compliance concerns. Finally, information technology can enhance community groups' ability to participate in governance efforts by conducting their own sampling, sharing results with EPA, and engaging regulated parties through both informal and formal mechanisms. Notwithstanding the transformative potential of new information technologies, "[t]he study of [information and communications technology] and its relationship to legal and regulatory systems is a topic that is still in its infancy as the subject of academic attention."[32]

This Article assesses the promises and pitfalls of relying on new technologies to generate and use new data sources (or increase the utility of existing sources) to improve environmental compliance and enforcement. In doing so, it identifies some important technical and practical challenges facing those, including government agencies, who seek to rely on these sources. The Article highlights the importance of addressing these challenges by reviewing an ongoing EPA initiative called Next Generation Compliance (or Next Gen), which aims to "transform" traditional environmental enforcement practices by relying more heavily on advanced monitoring and reporting technologies.[33] The fate of that effort is likely to be shaped by the extent to which the agency is able to exploit emerging technologies and recognize and respond effectively to challenges in doing so.

Part I of the Article describes the technological revolution that has enabled the generation and mining of new data streams that have the capacity to influence environmental compliance and enforcement. Part I also identifies a series of significant challenges in using this information to promote environmental compliance and enforcement. We categorize these challenges according to the activities that relate to the data—primarily, data collection and analysis.

USDA AGRIC. RESEARCH SERV., BIG DATA AND COMPUTING: BUILDING A VISION FOR ARS INFORMATION MANAGEMENT 7 (2013), https://www.ars.usda.gov/ARSUserFiles/20800500/BigDataReport_Mar-7-2013.pdf (discussing need to "[i]nvest in high-priority enhancements in scientific IT capabilities" to "[i]mprove data storage, computational resources, and network infrastructure").

32. Rónán Kennedy, *Rethinking Reflexive Law for the Information Age: Hybrid and Flexible Regulation by Disclosure*, 7 GEO. WASH. U. J. ENERGY & ENVTL. L. 124, 125 (2016). Professor Kennedy contends that "no coherent perspectives, approaches, or frameworks have developed" on the relationships between information technology and environmental regulation. *Id.*

33. Giles, *supra* note 19, at 26 ("As we continue to learn about ways to strengthen compliance, and take advantage of advances in technology, Next Gen can transform our protection work even in a time of declining budgets."). Greater reliance on information technologies, however, will also require agencies to commit resources, sometimes in significant amounts, to setting up and maintaining data collection and analysis programs.

In Part II, we analyze the ways in which the information technology revolution may influence environmental compliance and enforcement. Subpart A focuses on EPA's efforts through its Next Gen initiative to use new data sources and technologies. Subpart B covers the use of new information technologies by regulated entities. Subpart C deals with the rise of citizen science and its relationship to the emergence of new information sources of potential value in fostering increased environmental regulatory compliance. In reviewing this governance landscape, we identify some of the opportunities and challenges EPA is likely to face as it seeks to maximize the potential value and effective use of these technologies to improve compliance with the environmental laws.

I. THE INFORMATION TECHNOLOGY REVOLUTION: ITS POTENTIAL AND CHALLENGES FOR ENVIRONMENTAL COMPLIANCE AND ENFORCEMENT

The development of information technology in the last few decades has been hailed as having revolutionary impacts on society.[34] Computers and devices linked to them, such as sensors, are capable of producing data where it was once scarce[35] and in volumes that dwarf previously available information. These same technologies are capable of analyzing data more quickly and thoroughly than ever before. After reviewing some potential terminological questions, this Part identifies several challenges relating to data collection, transport, and analysis presented by the use of new data sources as a tool in enforcement-related decision making and highlights the need for EPA to prepare for and respond to them.

A. A Threshold Challenge: Defining the Key Terms

Advances in information technology have prompted a new vocabulary that includes terms such as "data mining" and "big data."[36] The use of these terms tends to be context specific, as a report to the President noted in 2014:

34. See Liane Colonna, *A Taxonomy and Classification of Data Mining*, 16 SMU SCI. & TECH. L. REV. 309, 314 (2013) (citing *10 Breakthrough Technologies*, MIT TECHNOLOGY REVIEW (2001), http://www2.technologyreview.com/tr10/?year=2001) (listing "data mining" as "one of the ten emerging technologies that would change the world"). Professor Liane Colonna lists some of the industries and activities most subject to change as a result of new information technologies, including the financial, health care, and telecommunications industries, education, sports, national security, and law enforcement. *Id.* at 351–66.

35. *See, e.g.*, Bennett B. Borden & Jason R. Baron, *Finding the Signal in the Noise: Information Governance, Analytics, and the Future of Legal Practice*, 20 RICH. J.L. & TECH. 7, 22 (2014) (quoting Kenneth Neil Cukier & Viktor Mayer-Schoenberger, *The Rise of Big Data: How It's Changing the Way We Think About the World*, COUNCIL ON FOREIGN RELATIONS (Apr. 3, 2013), https://www.foreignaffairs.com/articles/2013-04-03/rise-big-data) ("Big data is . . . characterized by the ability to render into data many aspects of the world that have never been quantified before; call it 'datafication.'").

36. Consideration of these issues has included book-length treatment and has also provided the basis for academic symposia. *See, e.g.*, Symposium, *Big Data Future Part One*, 10 I/S: J. L. & SOC'Y FOR INFO. SOC'Y 671 (2015); Symposium, *Big Data Future Part Two*, 11 I/S: J. L. & SOC'Y FOR INFO.

> There are many definitions of "big data" which may differ depending on whether you are a computer scientist, a financial analyst, or an entrepreneur pitching an idea to a venture capitalist. Most definitions reflect the growing technological ability to capture, aggregate, and process an ever-greater volume, velocity, and variety of data.[37]

A widely used definition centers on the three "Vs." It conceives of big data as "high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."[38] Big data is thus a term that has been used to refer to large volumes of information,[39] the techniques used to generate and

---

SOC'Y, 1 (2015). "The terms big data and big data analytics originally derive from the terms artificial intelligence, business intelligence, and business analytics; terms used in the 1950s, 1990s, and 2000s, respectively." Lieke Jetten & Stephen Sharon, *Selected Issues Concerning the Ethical Use of Big Data Health Analytics*, 72 WASH. & LEE L. REV. ONLINE 394, 395 (2016).

37. EXEC. OFFICE OF THE PRESIDENT, BIG DATA: SEIZING OPPORTUNITIES, PRESERVING VALUES 2 (May 2014), https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf [hereinafter SEIZING OPPORTUNITIES]. Some commentators have suggested a lack of consensus concerning the definition of big data. *See, e.g.*, Linda K. Breggin et al., *Big Data – Enabling Big Protection for the Environment*, *in* BIG DATA CHALLENGES IN EVIDENCE-BASED POLICYMAKING 113, 114 (H. Kumar Jayasuriya & Kathryn Ritscheske eds., 2015); *see also* Eytan Adar, *The Two Cultures and Big Data Research*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 765, 765 (2015) (referring to "[t]he lack of an agreed-upon definition for Big Data").

38. John Pavolotsky, *Privacy in the Age of Big Data*, 69 BUS. LAW. 217, 217 (2013) (quoting Svetlana Sicular, *Gartner's Big Data Definition Consists of Three Parts, Not to Be Confused with Three "V"s*, FORBES (Mar. 27, 2013, 8:00 AM), http://www.forbes.com/sites/gartnergroup/2013/03/27/gartners-big-data-definition-consists-of-three-parts-not-to-be-confused-with-three-vs/); *see also* Breggin & Amsalem, *supra* note 28, at 10,984 (citing *IT Glossary: Big Data*, GARTNER, INC., http://www.gartner.com/it-glossary/big-data/); SEIZING OPPORTUNITIES, *supra* note 37, at 4; Peter M. Shane, *Foreword: Big Data Future and the First Decade of an Interdisciplinary Journal*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 671, 675 (2015).

Some observers have identified additional "Vs," including veracity, see Margaret Hu, *Small Data Surveillance v. Big Data Cybersurveillance*, 42 PEPP. L. REV. 773, 795 n.59.(2015), and value, see Allen P. Grunes & Maurice E. Stucke, *No Mistake About It: The Important Role of Antitrust in the Era of Big Data*, 14 ANTITRUST SOURCE 1, 2 n.13 (2015) ("Value is a fourth V which is related to the increasing socioeconomic value to be obtained from the use of big data. It is the potential economic and social value that ultimately motivates the accumulation, processing and use of data.") (quoting ORG. FOR ECON. CO-OPERATION & DEV., SUPPORTING INVESTMENT IN KNOWLEDGE CAPITAL, GROWTH AND INNOVATION 12 (2013)).

39. *See* Nicolas P. Terry, *Protecting Patient Privacy in the Age of Big Data*, 81 UMKC L. REV. 385, 389 (2012) ("Not surprisingly, 'big' data is frequently defined in terms of its size. It even finds definition from what it is not ('datasets whose size is beyond the ability of typical database software tools to capture, manage, and analyze') and what it might be (vague estimates as to the petabytes and exabytes of information that are being captured)."); Sean Fahey, *The Democratization of Big Data*, 7 J. NAT'L SECURITY L. & POL'Y 325, 325 (2014) ("[O]ne can define big data as a collection of data that is so large that it exceeds one's capacity to process it in an acceptable amount of time with available tools."); ENVTL. L. INST., BIG DATA AND ENVIRONMENTAL PROTECTION: AN INITIAL SURVEY OF PUBLIC AND PRIVATE INITIATIVES 3 (2014), https://www.eli.org/sites/default/files/eli-pubs/big-data-and-environmental-protection.pdf ("'Big data' is commonly defined as data that are too large, created too quickly, or structured in such a manner as to be difficult to collect and process using traditional data management systems.") [hereinafter ELI, BIG DATA].

One difficulty with a volume-based approach is that it "incorporates a moving definition of how big a dataset needs to be in order to be considered big data . . . [because] . . . as technology advances over time, the size of datasets that qualify as big data will also increase." Roslyn Fuller, *Structuring Big* disseminate it, and the methods used to analyze it.[40] Groups like the U.S. National Institute of Standards and Technology, the International Standards Organization, and the World Wide Web Consortium[41] have made efforts to develop a common set of big data definitions, taxonomies, formats, and reference architectures.[42]

We do not seek here to provide a universally applicable definition of data mining, big data, or related information technology terms, or even a set of terms that will be appropriate for use in environmental law and policy contexts. Big data is only a part of "the newly emergent field of 'analytics,'" in which "data, statistical and quantitative analysis, explanatory and predictive models,[43] and fact-based management" are extensively used "to drive decisions and add value."[44] Rather, our purpose is to highlight some of the challenges an agency with a significant compliance promotion portfolio is likely to confront in taking advantage of this new capacity.

As Part II explains, EPA has or soon will have vast troves of new data at its disposal. The data are being generated by the agency itself, regulated entities, third-party auditors,[45] and broader civil society, including environmental nongovernmental organizations and community groups. But what can and will EPA do with all this data? In particular, can it collect, analyze, disseminate, and use the data in ways that enhance compliance with federal and state environmental regulatory duties? The promise of data

---

Data to Facilitate Democratic Participation in International Law, 42 INT'L J. LEGAL INFO. 504, 505 (2014) (quoting JAMES MANYIKA ET AL., BIG DATA: THE NEXT FRONTIER FOR INNOVATION, COMPETITION AND PRODUCTIVITY (McKinsey Global Inst., 2011)).

40. *See* Terry, *supra* note 39, at 391 (stating that "'big data' refers both to the ability to store and aggregate these giant datasets and the availability of increasingly powerful data mining and analysis techniques"); *see also* Daniel J. Solove, *Introduction: Privacy Self-Management and the Consent Dilemma*, 126 HARV. L. REV. 1879, 1889 (2013) (stating that "[m]odern data analytics . . . is also loosely referred to as data mining or 'Big Data'"); Breggin & Amsalem, *supra* note 28, at 10,985 ("The phrase 'big data' often is used to describe not only the data, but also the methods used to sift through and make sense of them, essentially making mountains of information useful."); *cf.* Neil M. Richards & Jonathan H. King, *Big Data Ethics*, 49 WAKE FOREST L. REV. 393, 394 (2014) (using the term "'big data' . . . to denote the collection and storage of large data sets" and "'big data analytics' . . . to denote inferences and predictions made from large data sets").

41. The Consortium is a standard-setting body for the World Wide Web. *See About W3C*, W3C, https://www.w3.org/Consortium/ (last visited Jan. 22, 2017).

42. Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 545–46 (2014).

43. *See* Anthony J. Casey & Anthony Niblett, *Self-Driving Laws*, 66 U. TORONTO L.J. 429, 431–32 (2016) ("New machine-learning techniques are outperforming traditional regression approaches to prediction."); Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards* 2–3 (U of Chicago, Public Law Working Paper No. 550 2015), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2693826 ("Innovations in big data and artificial intelligence will make it increasingly easy to predict the outcomes that certain behavior will produce.") [hereinafter Casey & Niblett, Death of Rules].

44. Borden & Baron, *supra* note 35, at 23 (quoting THOMAS H. DAVENPORT & JINHO KIM, KEEPING UP WITH THE QUANTS: YOUR GUIDE TO UNDERSTANDING AND USING ANALYTICS 3 (2013)).

45. *See* David A. Hindin & Jon D. Silberman, *Designing More Effective Rules and Permits*, 7 GEO. WASH. J. OF ENERGY & ENVTL. L. 103, 113 (2016) ("Third-party programs use independent entities to report information on regulated entities to the government or assess and verify whether the entities are meeting their regulatory obligations.").

analytics to foster higher levels of compliance and more effective enforcement is not unique to environmental law. As others have noted, "[b]ig data analytics can revolutionize law enforcement with its ability to . . . 'uncover hidden patterns, correlations, and other insights'"[46] both in individual criminal[47] and business regulatory contexts.[48]

B. Systemic Challenges in the Generation of Data and Use of Data Analytics

A shift to greater reliance on new or enhanced data streams and improved capacity to mine that data nevertheless poses challenges to EPA or any organization seeking to bolster compliance and enforcement by relying on them. This subpart identifies some of those challenges and explores how EPA might meet them in ways that maximize the value of new information technologies.

As discussed below, algorithmic and theoretical advances in machine learning are revolutionizing scientific discovery and spurring the creation of new technologies.[49] In the environmental policy arena, "[b]ig data analytics are increasingly being used to shed light on patterns and predict future trends, in an effort to understand business processes [and] support decisionmaking" in various regulatory contexts, including environmental enforcement.[50] For

46. Ismail Cem Kuru, *Your Hard Drive Is Almost Full: How Much Data Can the Fourth Amendment Hold*, 2016 U. ILL. J.L. TECH. & POL'Y 89, 92 (internal quotations omitted).

47. *See* Peter Segrist, *How the Rise of Big Data and Predictive Analytics Are Changing the Attorney's Duty of Competence*, 16 N.C. J.L. & TECH. 527, 568 (2015) ("[L]aw enforcement agencies are also applying big data analytics to identify specific individuals whom the data indicates warrant additional scrutiny."); Dennis D. Hirsch, *The Glass House Effect: Big Data, the New Oil, and the Power of Analogy*, 66 ME. L. REV. 373, 376 (2014) ("[C]ollecting massive amounts of surveillance camera data and mining it for law enforcement purposes . . . promises to reduce crime and increase personal safety."); David Gray et. al., *Fighting Cybercrime After* United States v. Jones, 103 J. CRIM. L. & CRIMINOLOGY 745, 798 (2013) (discussing the use of data analytics to prevent cybercrime).

48. *See* Joy Heath, *Government Highlights New Focus on Physician Fraud*, 27 HEALTH LAW 36, 38 n.1 (2015) (referring to the use of "data analytics and the combined resources of Federal, State, and local law enforcement entities to prevent and combat health care fraud, waste, and abuse"). The Securities and Exchange Commission (SEC) has created an entire division, the Division of Economic and Risk Analysis, "to integrate financial economics and rigorous data analytics into the core mission of the SEC. The Division is involved across the entire range of SEC activities, including policy-making, rule-making, enforcement, and examination." *About the Division of Economic and Risk Analysis*, SEC.GOV, https://www.sec.gov/dera/about (last visited Dec. 29, 2016); *see also* Professor Henry T. C. Hu, Allan Shivers Chair in the Law of Banking and Finance, University of Texas Law School, Keynote Address: The SEC, Dodd-Frank, and Modern Capital Markets, *in* 7 N.Y.U. J.L. & BUS. 427, 434–36 (2011) (describing the use of data analytics in securities regulation enforcement).

49. *See infra* notes 95-98 and accompanying text.

50. Breggin & Amsalem, *supra* note 28, at 10,986. *See also* Ann Klee, *The Digital Transformation of Environment, Health, and Safety*, 33 ENVTL. F. 17, 17 (characterizing data analysis as "the most transformative part of the new industrial revolution" resulting from new information technology); Juan Carlos Rodriguez, *EPA Enforcement Will Stay Tough Post-Obama, Giles Says*, LAW360 (Aug. 9, 2016, 6:55 PM), http://www.law360.com/articles/824039/epa-enforcement-will-stay-tough-post-obama-giles-says (quoting a top EPA enforcement official's prediction that data analytics is "going to grow exponentially in the coming years," increasing the agency's "ability to use data to find

example, agencies are using tools such as a custom database created by IBM and sophisticated analytics to map interrelated criminal activity. These tools allow agencies to draw connections between apparently unrelated cases involving illegal trafficking in hazardous substances or endangered species and share information with governments combatting eco-crimes.[51] Agencies are also using tools, such as online mapping programs, to provide information to environmental emergency responders and resource managers seeking to prepare for and coordinate responses to oil spills or other environmental disasters.[52]

The revolution in information technology has the potential to improve understanding of the state of environmental compliance. The combined efforts of government, regulated entities, and civil society will enable the generation and distribution of enormous quantities of new information about environmental conditions and performance. The revolution will provide new tools for analyzing the data that will enable public and private decision makers to adjust their practices in ways that improve compliance.

But revolutions do not always occur seamlessly. Instead, those participating in them often fail to identify or seize new opportunities. They take wrong turns and experience unintended consequences. Accordingly, efforts to incorporate newly available data generated by technological advances and new analytical techniques—including predictive algorithms—into environmental governance mechanisms, such as EPA's compliance and enforcement programs, are not likely to be smooth. To ease the transition, EPA should take steps to prepare for and respond to significant challenges presented by the use of new data sources as a tool in enforcement-related decision making. In this subpart, we identify several such challenges that relate to data collection, dissemination, and analysis.[53]

1. Data Collection, Storage, and Transport

The data collection challenges facing policy makers relying on new or enhanced data sources are myriad. Some of the most prominent challenges are likely to be

- gathering enough information and the right kinds of information;

---

serious problems, to identify criminal activity and to help us figure out where we should be focusing our time").

51. *See* Breggin & Amsalem, *supra* note 28, at 10,987; David Maxwell Braun, *Big Data Analytics Helping to Protect Big Cats*, NAT'L GEOGRAPHIC: CAT WATCH (Feb. 12, 2013), http://voices.nationalgeographic.com/2013/02/12/big-data-and-analytics-helping-to-protect-big-cats/.

52. *Environmental Response Management Application*, NOAA.GOV, http://response.restoration.noaa.gov/maps-and-spatial-data/environmental-response-management-application-erma (last visited Dec. 29, 2016).

53. The volume of the data available to an agency as a result of modern information technologies may itself present problems. *See, e.g.*, Kennedy, *supra* note 32, at 138 (noting that "it is possible to drown in data and for decisionmakers to be overwhelmed by the range of facts and figures that they must assimilate in order to come to a conclusion").

- weeding out poor quality or unreliable data resulting from poor equipment or human error resulting from lack of training in the use of equipment;
- aggregating information collected at different levels of temporal and spatial frequency;
- adjusting verification and quality control requirements to different intended uses of the data;
- protecting against hacking of computer systems that store the data resulting in data corruption; and
- providing data storage and transport systems that minimize errors and loss of data.[54]

The strength of policy-based decisions depends on the quality of the data on which decision makers rely.[55] As one information technology services expert put it, "[w]hile data is a catalyst for innovation, data governance is a catalyst for quality, and value is derived from well-governed quality data."[56] Poor data quality can adversely affect decisions by individuals, firms, and

---

54. *See* Lise Getoor et al., *Computing Research and the Emerging Field of Data Science*, COMPUTING RES. NEWS, (Oct. 2016), http://cra.org/crn/2016/10/computing-research-emerging-field-data-science/ (contending that "[f]rom a data management perspective, data science requires a much deeper understanding and representation of how data is acquired, stored and accessed," that "very large data volumes, very high data rates, and very large numbers of users, demand new systems and new algorithms," that "many classic statistical assumptions and machine learning techniques do not fit current data science needs," and that "challenges in scale and heterogeneity also fundamentally change how users interact with data and models, how the data is visualized, what algorithms are needed to support understanding and interpretation of the results of data science models, how decisions are made, and how user feedback is acquired and incorporated"); W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine*, 37 CARDOZO L. REV. 1401, 1412 (2016) (stating that, in using big data for medical treatment, "firms must gain access to the substantial amounts of data in electronic form"). These do not exhaust the range of challenges likely to be presented by greater reliance on big data, but we think they are among the most important and likely to recur. Others, which are beyond the scope of this Article, include privacy and national security concerns. *See supra* notes 24, 27 and accompanying text. For discussion of some of the legal and policy challenges, see Mattioli, *supra* note 42, at 536 ("The nondisclosure of data's provenance and pedigree . . . [can] impede[] data reuse, which in turn can prevent innovative applications of the big data method."); Wagner & Finkelman, *supra* note 23, at 599 (discussing security and privacy challenges); Margaret Hu, *Big Data Blacklisting*, 67 FLA. L. REV. 1735, 1735 (2015) (discussing constitutional challenges); Hu, *supra* note 38, at 785 (discussing "challenges of big data-driven national security policymaking and the role of big data cybersurveillance in national security law").

55. "[P]roblems can arise when data are incorrect or outdated, even if there are large quantities of it. This is often summed up by the adage 'garbage in, garbage out.'" Breggin & Amsalem, *supra* note 28, at 10,992.

56. Barbara L. Cohn, *Data Governance: A Quality Imperative in the Era of Big Data, Open Data, and Beyond*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 811, 811 (2015). Dr. Barbara Cohn defines data governance as "a framework which formalizes the roles, functions, and procedures within which an organization's data is well managed and enabled as a strategic asset." *Id.* at 813. She lists the core elements of effective data governance as leadership, adaptability, structure, standards, and objectives. *Id.* at 815; *cf.* Brian H. Cameron, *The Need for Enterprise Architecture for Enterprise-Wide Big Data*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 827, 845 (2015) ("Similar to other IT projects, it is necessary for enterprises to define what the outcomes of the big data project will be, who will benefit from it, and how they will benefit. Hence, as long as big data projects are considered to pose purely technical issues, the failures will continue to pile up.").

governments.[57] EPA has experienced a long history of data accuracy problems in its generation and use of compliance-related data.[58] Erroneous or inaccurate data may prevent data analysts from drawing useful insights into the nature of the problems being investigated, such as adverse ambient environmental conditions that may be linked to undiscovered noncompliance. Errors or inaccuracies may also hinder discovering the best solutions to address those problems.[59] As the discussion below indicates, even a shift to advanced information-gathering technologies, such as electronic reporting, is not likely to eliminate data quality problems, though it may alleviate them.[60]

Data quality problems are likely to include both incomplete and inaccurate data.[61] Large information-gathering efforts, such as the ones EPA envisions as the foundation of its Next Gen efforts to enhance compliance and enforcement, often depend on the aggregation of information derived from multiple sources. If EPA relies on those outside the agency to supply it with the data it uses to drive a transformation of its compliance and enforcement program, data gaps may develop as a result of the lack of an integrated, systematic approach to data collection. Data generated by individuals and community groups may be "self-selected with unsure representativeness,"[62] although some community groups have sought to monitor in areas traditionally neglected by government monitors.[63] Even if those outside the agency are collecting the right kinds of information and doing so without introducing significant inaccuracies, transmission to EPA of information collected by states, community groups, or third-party auditors (or to state enforcement officials by communities and third parties) may be delayed.

---

57. James T. Graves et. al., *Big Data and Bad Data: On the Sensitivity of Security Policy to Imperfect Information*, 83 U. CHI. L. REV. 117, 121 (2016). As Professor Kennedy has explained:

> The provision of information, by itself, is not a form of risk assessment. We should not assume that simply because information is publicly available, it is accurate, properly understood, or complete. Analysis of the [Toxic Release Inventory under the Emergency Planning and Community-Right-to-Know Act, 42 U.S.C. § 11023 (2012)] data has revealed that it has contained significant errors in recording the quantity and location of toxic releases. If it is not carefully designed, [an environmental regulatory program relying on information disclosure] will contain many of the weaknesses ascribed to command-and-control environment regulation—an unwarranted focus on major sources, a lack of discrimination between pollution types, or little incentive for further research.

Kennedy, *supra* note 32, at 136.

58. *See, e.g.*, *Dynamic Governance, Part I*, *supra* note 29, at 586–89; Markell & Glicksman, *A Holistic Look*, *supra* note 29, at 47–48 (describing inaccuracy and incompleteness of data on compliance and enforcement).

59. *See* Price, *supra* note 54, at 1414.

60. *See infra* Part I.B.2.

61. *See* Graves et. al., *supra* note 57, at 121–31 (reviewing data quality problems in developing national security policies).

62. Harvey J. Miller, *Space-Time Data Science for a Speedy World*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 705, 712 (2015).

63. Gregg P. Macey, *The Architecture of Ignorance*, 2013 UTAH L. REV. 1627, 1659 (2013) ("The public has begun to question the spatial location of data, taking samples on residential streets and in schoolyards at ground level, places ignored by government stations.").

Even if these entities provide information to EPA, the diverse nature of the data sources increases the challenge of assuring that it is accurate, of high quality, and relevant to the uses to which the agency wants to put it. The usefulness of information, such as environmental monitoring data, will depend in part on the quality of the devices used to generate it. The reliability of data generated by individuals and community groups may warrant special attention because of the types of monitoring equipment involved and the possibility that those supplying the information lack the training to operate the equipment properly. The kinds of low-cost sensors that tend to be used in these data collection efforts may not be as accurate as the kinds of monitoring devices traditionally used by regulators and regulated entities, and may give rise to "false alarms."[64] In addition, the devices must be properly calibrated and operated in a fashion that is not likely to taint or otherwise render unhelpful the data they produce.[65] Calibration is critical to environmental monitoring. As EPA explained:

> Calibration is the process of checking and adjusting an instrument's measurements to ensure that it is reporting accurate data. Calibration compares the response of the instrument to a known reference value. Calibration is important because sensor performance can change over time. If at all possible, sensors should be calibrated for their response before, during, and after a set of data collections.[66]

The challenge of ensuring proper calibration will obviously be much greater if the devices generating enforcement-related data are being operated by myriad nongovernmental sources. Lack of sophistication, experience, and training in using the equipment used to collect data also are likely to produce transmission errors.

Data, such as environmental monitoring information, may be used for different purposes. These include ascertaining the need for more stringent emission control standards, devising strategies to improve ambient conditions, and identifying noncompliance and collecting evidence to support enforcement

---

64. *See* Douglas Main, *Your Office Air Is Killing You*, NEWSWEEK (June 2, 2016, 6:10 AM), http://www.newsweek.com/2016/06/10/indoor-air-pollution-revolution-465531.html; Patrick Ambrosio, *Low-Cost Air Monitoring Research Funded by EPA*, 47 ENV'T REP. (BNA) 2380 (2016) (noting that while inexpensive devices for monitoring air pollution are increasingly available, many "have not yet been widely tested"). For further discussion of the falling cost of certain kinds of monitoring devices, see *infra* Part II.C.1.

65. *See, e.g.*, Hindin & Silberman, *supra* note 45, at 111 (explaining that immediate feedback technology, which supplies regulated entities with ongoing alerts as to compliance status, "must be constructed to appropriate specifications and properly installed and calibrated to applicable standards to ensure their results are accurate and reliable").

66. RON WILLIAMS ET AL., EPA, AIR SENSOR GUIDEBOOK 21 (2014), https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=277996&simpleSearch=1&searchAll=air+sensor+guidebook. EPA has also published a report on the capacity of low-cost devices to measure volatile organic compound concentrations. *See generally* RON WILLIAMS ET AL., EPA, NEXT GENERATION AIR MONITORING (NGAM) VOC SENSOR EVALUATION REPORT (2015), https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=308114&simpleSearch=1&searchAll=Next+generation+air+monitoring.

action. As a result, regulators will likely need to establish different verification and quality control requirements based on the intended use of the data. As researchers at the Environmental Law Institute have pointed out, "[t]ypically, the more regulatory or enforcement-oriented the goal may be, the more detailed or prescriptive are the legal requirements."[67]

In the enforcement context, if an agency is collecting data to identify or target noncompliance, it will need to ensure that the data relate to the variables being monitored for compliance. Further, the data must be collected at appropriate locations. For example, as the discussion in Part II indicates, EPA has begun expanding the geographic scope of environmental monitoring by generating (or requiring regulated entities to generate) information on ambient conditions at facility fencelines.[68] Because EPA intends to use such monitoring to help identify noncompliance, it must design data gathering efforts with that objective in mind.[69]

Data collection therefore requires coordination among multiple data sources,[70] checks on quality, and compliance with legal and regulatory requirements.[71] Some see these data collection and transmission challenges as

---

67. Daniel Rejeski & James McElfish, *Citizen Science*, 33 ENVTL. F. 62, 63 (2016). The Virginia Department of Environmental Quality intends to use citizen-generated data "to educate the community, to assist local governments in land use planning, to supplement data for university and professional studies, and to assist local soil and water conservation districts in prioritizing watershed work for best management practices." *Levels of Citizen Water Quality Data in Virginia*, VA. DEP'T OF ENVTL. QUALITY, http://www.deq.virginia.gov/ (last visited Dec. 30, 2016) [hereinafter *Citizen Water Quality Data*].

68. Fenceline monitoring is not an entirely new component of environmental regulation. *See, e.g.*, Thomas O. McGarity, *Hazardous Air Pollutants, Migrating Hot Spots, and the Prospect of Data-Driven Regulation of Complex Industrial Complexes*, 86 TEX. L. REV. 1445, 1470–73 (2008) (discussing the use of fenceline monitoring in Texas's regulation of hazardous air pollutant emissions).

69. Macey, *supra* note 63, at 1659.

70. *See* Matthew Gordon, *Big Data: It's Not the Size That Matters*, 7 J. NAT'L SECURITY L. & POL'Y 311, 313 (2014) (observing that big data are sometimes "pushed into databases with only rudimentary user interfaces, and data spread across multiple incompatible databases can't be combined or compared"); *see also* McGarity, *supra* note 68, at 1481–83 (discussing limitations on the accuracy of modern, mobile monitoring technologies).

71. Price II, *supra* note 54, at 1415. Matthew Gordon provides an example:

> The goal of data integration should be to provide not only a mechanism for importing and normalizing data from multiple sources, but also a framework for combining both structured and unstructured data together on the same continuum. A simple but powerful example is ferreting out insider trading. Such investigations may rely on trading records from a spreadsheet, phone records from a database, e-mails from an enterprise IT system, and company earnings announcements from the internet. None of these can demonstrate insider trading conclusively, but taken together, they can paint a very compelling picture.

Gordon, *supra* note 70, at 314–15. Gordon refers to data integration as one of the "Four Pillars" of effective use of data, along with search and discovery, knowledge management, and collaboration. *Id.* at 314. For an example of the difficulties involved in integrating multiple datasets, see Michael Batty, *Does Big Data Lead to Smarter Cities? Problems, Pitfalls and Opportunities*, 11 I/S: J.L. & POL'Y FOR INFO. SOC'Y 127, 139–44 (2015) (concerning travel on the London underground system).

an intractable problem, contending that relatively few analytics solutions "work robustly with multimodal and heterogeneous data types."[72]

But even if the data collected and transmitted to EPA are initially of high quality, government databases may be susceptible to hacking by outsiders. Hacking could result in disclosure of information the government regards as confidential[73] or corruption of stored data that impairs its utility in supporting enforcement action.[74] This problem is obviously not confined to databases comprised of information supplied by those outside the government.

EPA's development of protocols for the generation and collection of data will be critical to its ability to achieve Next Gen goals. Both researchers and policy makers have grappled with these technical challenges. As noted above, several groups have undertaken efforts to develop common definitions and other key elements of protocols.[75] Environmental agencies have also made efforts to develop quality control protocols. Virginia's Department of Environmental Quality, for example, has developed three levels of data quality for citizen-monitoring efforts based on the level of data quality and the authorized uses of the data provided to the agency. Among other things, it anticipates that these data will be useful to it in identifying those water bodies for which future agency monitoring is most critical.[76] As one observer has noted, "providing data standards . . . is essential for meaningful data exchanges, which is a critical part of transparency and accountability."[77]

2. Analysis

Data collection is merely the first step in the process of using new and existing information sources to improve environmental policy actions. The next step is analysis of the data collected. EPA has long faced challenges in analyzing its data efficiently and effectively and in using the resulting

72. Ashit Talukder, *Big Data Open Standards and Benchmarking to Foster Innovation*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 799, 802 (2015).

73. *See, e.g.*, Scott R. Peppet, *Regulating the Internet of Things: First Steps Toward Managing Discrimination, Privacy, Security, and Consent*, 93 TEX. L. REV. 85, 117 (2014) (discussing "the vulnerability of these consumer devices to hacking and other security breaches").

74. *See* Chad Squitieri, Note, *Confronting Big Data: Applying the Confrontation Clause to Government Data Collection*, 101 VA. L. REV. 2011, 2027–28 (2015) (discussing "[d]ata, including data stored in the 'cloud,' is susceptible to corruption while in storage. . . . Stored data is also susceptible to destruction"); Mark Rappa, Executive Director, Institute for Advanced Analytics, N.C. State Univ., Symposium Remarks: The Evolving Role of the Corporate Counsel: How Information Technology Is Reinventing Legal Practice (2013), *in* 36 CAMPBELL L. REV. 383, 398 (2013) ("We also have to worry about data corruption. . . . There are people with mal intent who are very interested in corrupting data.").

75. *See supra* note 42 and accompanying text.

76. *Citizen Water Quality Data*, *supra* note 67; *see also* ENVTL. LAW INST., CLEARING THE PATH: CITIZEN SCIENCE AND PUBLIC DECISION MAKING IN THE UNITED STATES 1 (2016), http://www.eli.org/sites/default/files/eli-pubs/clearing-path-eli-report.pdf (suggesting "appropriate design considerations for projects to clear the path toward greater governmental access to, and reliance on, citizen science") [hereinafter ELI, CLEARING THE PATH].

77. Gary D. Bass, *Big Data and Government Accountability: An Agenda for the Future*, 11 I/S: J.L. & POL'Y FOR INFO. SOC'Y 13, 32–33 (2015).

analysis.[78] Increasing volumes of data, from an increasing variety of sources, collected in an increasing variety of ways, will inevitably create both challenges and new opportunities in the agency's efforts to interpret the significance of that data.[79] A potentially daunting challenge for users of big data is designing methods for analyzing the data collected because "without the analytic ability to unlock key information and patterns, big data sets are of limited use."[80]

The three "Vs" that often characterize big data all present potential interpretive problems.[81] The data streams produced by new information technology can arrive quickly, in complex formats, and from a variety of sources. This mélange of information contributes to "the challenge of finding signals in the noise"[82] and thus, detecting patterns or relationships that are useful in making the decisions that the collection of data is supposed to assist. As one observer put it, the difficulty in *processing* big data "can be a result of the data's volume (e.g., its size as measured in petabytes), its velocity (e.g., the number of new data elements added each second), or its variety (e.g., the mix of different types of data including structured and unstructured text, images, videos, etc. . . .)."[83] Another analogized the analysis of big data to finding a needle in a haystack.[84]

Government officials responsible for fostering the use of new technologies with the potential to support U.S. economic growth and competitiveness have recognized the analytical conundrums that those technologies may present. The Program Director for Data Science at the National Institute of Standards and

78. *See* Joanna Lau, Comment, *Nothing but Unconditional Love for Conditional Registrations: The Conditional Registration Loophole in the Federal Insecticide, Fungicide, and Rodenticide Act*, 44 ENVTL. L. 1177, 1196 (2014) (referring to EPA's "inefficient data-tracking methods"); Carol S. Curme, *Regulation of Pesticide Residues in Foods: Proposed Solutions to Current Inadequacies Under FFDCA and FIFRA*, 49 FOOD & DRUG L.J. 609, 620 (1994) (referring to EPA's "inefficient data collection").

79. *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 3, 14–15 (2015) (arguing that while the large amounts of data collected by internet and financial services companies may create greater efficiency through predictability, secrecy in how these corporations use data creates problems of manipulation and accountability); *cf.* Andrea Roth, *Trial by Machine*, 104 GEO. L.J. 1245, 1269 (2016) ("Many crime-detecting gadgets and software tend to be shrouded in 'inscrutable black box[es]' that 'hide the workings' in shiny steel contraptions or computer code.").

80. Breggin & Amsalem, *supra* note 28, at 10,986. *See also* Kristin Madison, *Health Regulators as Data Stewards*, 92 N.C. L. REV. 1605, 1608–09 (2014) ("While amassing data can be an important first step in generating the information critical for [policy decisions], these data need to be analyzed and distilled before they can be used effectively by . . . stakeholders" in areas such as health care policy.).

81. For discussion of the three Vs, see *supra* note 38 and accompanying text.

82. Borden & Baron, *supra* note 35, at 21.

83. Fahey, *supra* note 39, at 325.

84. Mattioli, *supra* note 42, at 557–58. Despite the challenges, techniques to find such needles exist and have been used with considerable success. *See generally* Yann LeCun et al., *Deep Learning*, 521 NATURE 436 (2015), http://www.nature.com/nature/journal/v521/n7553/full/nature14539.html (describing the use of "deep learning" to promote problem solving in speech recognition, visual object recognition, object detection, and other areas such as drug discovery and genomics).

Technology[85] has identified a series of challenges in understanding and measuring big data analytics solutions, though the scope of these challenges is likely subject to debate and dependent on context. These include ongoing efforts to improve evaluation methods, tools, and reference data; increase understanding of the usability of big data systems and solutions and of how the quality and context of input data affect derived conclusions; and the need to determine how to evaluate which components are best suited to specific families of tasks.[86] The challenges identified by the Program Director may be among the types of questions that EPA will encounter and need to address if the promise of increased volumes of data and enhanced analytical capabilities as a compliance-enhancement tool is to be realized.[87]

One interpretive problem relevant to the environmental enforcement context involves connecting problematic environmental conditions to particular sources suspected of violating emission limits. The specific issue here is whether one can draw a causal link with sufficient statistical support between an observation, such as excessive chemical concentration at a particular location and time, and a violation attributable to a particular source.[88] This

85. Congress created the National Institute of Standards and Technology, which is currently housed within the Department of Commerce, in 1901 to improve the country's measurement infrastructure. According to the agency, "from the smart electric power grid and electronic health records to atomic clocks, advanced nanomaterials, and computer chips, innumerable products and services rely in some way on technology, measurement, and standards provided by the National Institute of Standards and Technology." *About NIST*, NAT'L INST. FOR STANDARDS & TIME, https://www.nist.gov/about-nist (last visited Dec. 30, 2016).

86. Talukder, *supra* note 72, at 805–06. For a general discussion of the promise and potential challenges presented by machine learning, see CHRISTOPHER BISHOP, PATTERN RECOGNITION AND MACHINE LEARNING (Springer 2007).

87. Statistical analysis of data uses samples to generalize about a larger population. *See* Sean Brian, Comment, *The Unexamined Life in the Era of Big Data: Toward A UDAAP for Data*, 40 U. DAYTON L. REV. 181, 183–84 (2015) ("More data lends greater predictive power to the generalizations that result from the study. . . . Rather than programming the proper response to every problem an application might encounter, machine learning allows a computer program to gather data until it learns how to respond."). "'Machine learning' refers to a subfield of computer science concerned with computer programs that are able to learn from experience and thus improve their performance over time. . . . [T]he idea that the computers are 'learning' is largely a metaphor. . . . Rather, [they learn] in a functional sense: they are capable of changing their behavior to enhance their performance on some task through experience." Harry Surden, *Machine Learning and Law*, 89 WASH. L. REV. 87, 89 (2014). Insight on the derivation of the term "machine learning" is provided by the statement that "[a]rtificially intelligent machines find 'hidden' or 'deep' connections in unstructured data to provide stronger predictions. In some sense, these machines are capable of 'learning.' They update to take into account whether their best guesses are correct or not." Casey & Niblett, Death of Rules, *supra* note 43, at 30. But the amount of data available to decision makers may itself pose analytical challenges. *See, e.g.*, Farnam Jahanian, *The Policy Infrastructure for Big Data: From Data to Knowledge to Action*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 865, 872–73 (2015); *see also* Bass, *supra* note 77, at 33 ("Sometimes the best way to hide key information is to bury it in massive datasets.").

88. The problem has been described with respect to a different environmental policy context as follows:

> In the era of data deluge, we are confronted with largescale time series data, i.e., a sequence of observations of concerned variables over a period of time. . . . A major data mining task for time series data is to uncover the temporal causal relationship among the time series. For

problem, which has persisted for decades, influenced the design of the Clean Water Act (CWA) in 1972. In light of the difficulty in identifying cause-and-effect relationships between discharges and ambient water quality, Congress instead chose to engraft a technology-based program for controlling point source discharges to surface water onto the pre-existing water quality-based regime. The technology-based approach to regulation bypassed in part the difficulty of drawing that causal connection based on the available scientific information, which had previously hindered enforcement efforts by the states.[89]

The use of more sophisticated monitoring devices, coupled with computer-driven analysis of the data generated, has the potential to identify those kinds of causal connections more easily than before.[90] But new data streams do not automatically make the desired causal connections or produce the information needed to support policy decisions. Nor do they necessarily provide the evidentiary foundations for enforcement actions.

For example, imagine an agency detects chemical concentrations that are of concern at location A at time t_a. Assume, too, that there are multiple polluters in the relevant vicinity. The violation may trigger an investigation into potential emission violations by nearby sources. However, successful enforcement action may require disaggregation of the observed chemical concentration to help reveal the chemical concentration emitted from location B at time t_b and the chemical concentration emitted from location C at time t_c. Although advanced information technologies may help to diminish the challenge of working backwards from problematic ambient conditions to responsible sources, agencies must determine how best to collect and analyze new data streams to enable them to do so.

Another analytical challenge is "ensur[ing] that causal inferences are not distorted by systematic biases. Analysts and users of research data must be familiar with the risks of selection bias, confounding bias, and measurement

---

> example, in the climatology, we want to identify the factors that impact the climate patterns of certain regions. . . . Developing effective and scalable data mining algorithms to uncover temporal dependency structures between time series and reveal insights from data has become a key problem in machine learning and data mining.

Mohammad Taha Bahadori & Yan Liu, *An Examination of Practical Granger Causality Inference*, PROCEEDINGS OF THE 2013 SIAM INT'L CONF. ON DATA MINING 467 (2013), http://epubs.siam.org/doi/pdf/10.1137/1.9781611972832.52. *See also* McGarity, *supra* note 68, at 1483 ("Once the mobile-monitoring team has identified a toxic hot spot, the agency must still isolate the source or sources of the emissions that caused the elevated concentrations before it can fully assess the nature of the residual risks posed by those sources and induce the responsible companies to take additional steps . . . to reduce those emissions.").

89. Robert L. Glicksman & Mathew R. Batzel, *Science, Politics, Law and the Arc of the Clean Water Act: The Role of Assumptions in the Adoption of a Pollution Control Landmark*, 32 WASH. U. J.L. & POL'Y 99, 118–19 (2010).

90. *See* Kennedy, *supra* note 32, at 126 ("Hitherto invisible environmental problems, such as the depletion of fish stocks, can be brought to light through analysis of data. The impact of emissions over time and at a distance can be better understood. The interconnection of environmental hazards, such as the composition and sources of polluted air, can be more easily tracked.").

bias."[91] A recent report to the President prepared by a panel that included two cabinet secretaries and the Directors of the Office of Science and Technology Policy and the National Economic Council cautioned that "[f]inding a correlation with big data techniques may not be an appropriate basis for predicting outcomes or behavior, or rendering judgments on individuals. In big data, as with all data, interpretation is always important."[92] Similarly, a geographer at the Ohio State University who specializes in spatial analysis and geocomputation has argued that "massive amounts of streaming data favor correlations over causality since the former can be derived quickly and easily while the later requires deliberate theorizing and testing."[93] Thus, the analytical challenges that big data users, such as EPA, face include "avoid[ing] pitfalls such as taking inappropriate actions based on correlated data that has no causal connection," and using advanced analytics "to improve understanding of causation in a regulatory context. . . ."[94]

The literature on machine learning, which includes a subfield on causality, provides new techniques for addressing such causation questions.[95] Machine learning is a cutting-edge research area at the interface of computer science and statistics, focused on developing algorithms for data analysis, with formal theoretical underpinnings.[96] Algorithmic and theoretical advances in machine learning are crucial to advances in big data analytics. Machine learning has revolutionized scientific discovery, and spawned new technologies.[97] Use of

91. Sharona Hoffman & Andy Podgurski, *Big Bad Data: Law, Public Health, and Biomedical Databases*, 41 J.L. MED. & ETHICS 56, 57 (2013).

> Selection bias can occur when analysts unknowingly employ a study group that is not representative of the population of interest. . . . Confounding bias is a systematic error that occurs because there exists a common cause of the treatment/exposure variable and the outcome variable. . . . Measurement biases are generated by errors in measurement and data collection resulting from faulty equipment or software or from human error.

Id. at 58; *see also* Breggin & Amsalem, *supra* note 28, at 10,991 ("As big data sets are increasingly used to foster environmental protection efforts, it will be important to recognize, plan for, and address potential pitfalls. Possible pitfalls include biased data collection, analysis, and interpretation and reliance on low-quality data."). Ultimately, "courts may have to address questions such as whether the data the agency relied upon is biased or was interpreted in a biased manner." *Id.* at 10,994.

92. SEIZING OPPORTUNITIES, *supra* note 37, at 7; *see also* Breggin & Amsalem, *supra* note 28, at 10,994 (stating that "correlation does not necessarily indicate a cause-and-effect relationship").

93. Miller, *supra* note 62, at 714.

94. Breggin et. al., *supra* note 37, at 121.

95. *See* Susan Athey & Guido Imbens, *Machine Learning Methods for Causal Effects*, NASONLINE.ORG, http://www.nasonline.org/programs/sackler-colloquia/documents/athey.pdf (last visited Dec. 30, 2016); Peter Spirtes, *Introduction to Causal Inference*, 11 J. MACHINE LEARNING RES. 1643, 1643 (2010).

96. *See generally* Mireille Hildebrandt, *Law as Information in the Era of Data-Driven Agency*, 79 MODERN L. REV. 1 (2016) (discussing the mathematical theory of information that informs computer systems); N. D. Lawrence, *Probabilistic Non-Linear Principal Component Analysis with Gaussian Process Latent Variable Models*, 6 J. OF MACHINE LEARNING RES. 1783, 1783–84 (2005) (discussing types of machine learning).

97. *See* Pedro Larrañaga et al., *Machine Learning in Bioinformatics*, 7 BRIEFINGS IN BIOINFORMATICS 86, 86 (2005); Peter van der Graf, *How Search Engines Use Machine Learning for Pattern Detection*, SEARCH ENGINE WATCH (Dec. 1, 2011), https://searchenginewatch.com/sew/

machine learning techniques has driven discoveries that promise increased understanding of problems, such as climate change, and more informed solutions to those problems.[98] But increased reliance on machine learning has not yet eliminated all of the pitfalls that efforts to analyze new data streams may pose.

For example, databases often have systematic biases. An example from the environmental world is the higher ambient temperature that exists in urban areas and near buildings.[99] As a result, sampling close to buildings would yield predictions that temperatures are higher than one would anticipate is the reality; and sampling that leaves out urban areas would yield the opposite result. To avoid an inadvertent bias, an agency like EPA has to be sure that the data is a representative sample of the distribution it is interested in learning about.

EPA has taken steps to address some of these analytical challenges, such as causal attribution problems. The fenceline monitoring requirements included in the agency's recently adopted regulations to control benzene emissions from petroleum refineries are an effort to do so, for example, by providing guidance on how to subtract from fenceline benzene concentration measurements background concentrations and amounts emitted by non-refinery sources or caused by fugitive emissions.[100] EPA, states, community groups, nongovernmental organizations, and individuals will generate monitoring data with the aim of identifying facilities likely to be in noncompliance based on their proximity to problematic ambient conditions. The results may be used to identify those facilities for which further monitoring of compliance status is

---

opinion/2129359/search-engines-machine-learning-pattern-detection; Cade Metz, *AI Is Transforming Google Search. The Rest of the Web Is Next*, WIRED (Feb. 4, 2016, 7:00 AM), https://www.wired.com/2016/02/ai-is-changing-the-technology-behind-google-searches/.

98. *See* M.J. Smith et al., *The Climate Dependence of the Terrestrial Carbon Cycle, Including Parameter and Structural Uncertainties*, 10 BIOGEOSCIENCES 583, 583 (2013) (discussing computational models for assessing the feedback between climate and the terrestrial carbon cycle); Yi Deng & Imme Ebert-Uphoff, *Weakening of Atmospheric Information Flow in a Warming Climate in the Community Climate System Model*, 41 GEOPHYSICAL RES. LETTERS 193 (2014) (finding that storm tracks are moving northward with climate change); Lindene Patton, *Advances in Attribution Science, Emergence of Aggressive Climate Litigation Changing the Landscape for Voluntary Disclosure Programs*, 47 ENV'T REP. (BNA) 2639 (2016) (discussing advances in attribution science that enhance scientists' ability to link extreme weather events to climate change); Saket Navlakha & Ziv Bar-Joseph, *Distributed Information Processing in Biological and Computational Systems*, 58 COMMS. ASS'N COMPUTING MACHINERY 94 (2015) (discussing use of new technologies and computational methods to study and model biological systems). For a list of (and links to) "data-driven" research projects concerning climate change funded by the National Science Foundation, see *Expeditions in Computing: Understanding Climate Change, A Data-Driven Approach*, UNIV. OF MINN., http://climatechange.cs.umn.edu/publications.php (last visited Jan. 1, 2017).

99. *See* Jessica E. Fliegelman, *The Next Generation of Greenwash: Diminishing Consumer Confusion Through A National Eco-Labeling Program*, 37 FORDHAM URB. L.J. 1001, 1004 (2010) ("[H]eat trapping in urban landscapes with buildings and pavement creates threats of rising temperatures . . . .").

100. Petroleum Refinery Sector Risk and Technology Review and New Source Performance Standards, 80 Fed. Reg. 75,178, 75,192 (Dec. 1, 2015) (to be codified at 40 C.F.R. pt. 60 & 63) [hereinafter Refinery Risk].

appropriate. Notwithstanding EPA's efforts to date to address causal attribution problems, the challenges posed by efforts to do so are likely to be ongoing.[101]

This Part has explored some of the data collection and analysis challenges users face in relying on data streams that have become available as a result of the development of new information technologies and data analytics methodologies. EPA is aware of those challenges as it seeks to transform its compliance and enforcement apparatus by taking advantage of these new technologies and analytical methods. The revolution in information technology will affect not only government enforcers, but also regulated entities and components of civil society interested in reducing threats to health and the environment. The next Part discusses some of the impacts of innovations in information technology on both governmental and nongovernmental actors. It does so through an exploration of the manner in which EPA's latest effort to transform its compliance and enforcement programs, Next Gen, seeks to enhance the capacity of all these actors to take full advantage of new information technologies as compliance enhancement tools.

II. Potential Uses of Technological Innovations in Environmental Enforcement

Both legal scholars and experts in computer science and informatics agree that the information supplied by newly available technologies is likely to spur

101. In addition to the challenges described above, environmental agencies seeking to take advantage of enhanced data collection and analysis tools must have the know-how and the resources to use them. *See* Roslyn Fuller, *Structuring Big Data to Facilitate Democratic Participation in International Law*, 42 INT'L J. LEGAL INFO. 504, 505 (2014) ("[B]y its very nature, big data can only be usefully exploited by those entities with access to the necessary processing tools to capture and assemble it—that is governments and/or corporations with large IT expenditures.").

In the private sector, firms often rely on technology startups that specialize in sifting through huge volumes of data. *See* Mattioli, *supra* note 42, at 558. Effective use of big data requires the expertise of what are called "'data scientists': people with skill sets that span computer science, statistics and business analysis." Jane Griffin, *Managing Disruptive Technologies in the Cloud*, BASELINE (Mar. 3, 2016), http://www.baselinemag.com/cloud-computing/managing-disruptive-technologies-in-the-cloud. The development of predictive algorithms, for example, requires "substantial time, programming experience, and computational resources." Price, *supra* note 54, at 1415–16. Among other things, the algorithms used to analyze the data must be validated. *Id.* at 1416. If government agencies such as EPA want to base policy decisions on the kinds of information produced by new information technologies, they, too, must develop or contract with others who already have such expertise. The budgetary constraints that have affected agencies such as EPA may limit the funds available to purchase necessary hardware or software or to hire or contract with experts capable of putting big data to good use. Observers have questioned whether other government agencies seeking to increase the use of big data have sufficient resources to do so. *See, e.g.*, Hoffman & Podgurski, *supra* note 91, at 58. Others have noted a "talent shortage, from deep analytical talent and supporting engineers, to big-data-savvy professionals." Angela Byers, *Big Data, Big Economic Impact?*, 10 I/S: J.L. & POL'Y FOR INFO. SOC'Y 757, 762 (2015). In a market in which demand exceeds supply, these professionals may gravitate to high-paying jobs in the private sector rather than work in the government. *See* CARY COGLIANESE, INST. FOR L. & ECON., UNIV. PENN. L. SCH., OPTIMIZING GOVERNMENT FOR AN OPTIMIZING ECONOMY 7 (2016), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2789690 ("[T]he federal government's information technology infrastructure needs to rise to the task.").

important innovations.[102] In the environmental enforcement context, these may include ways to identify noncompliance, prioritize use of government investigation and enforcement resources, and engage regulated entities and affected communities in seeking ways to improve compliance performance. In the environmental law and policy arena, newly available information streams are already being used to "support a range of core government functions, including priority setting,[103] enforcement and compliance,[104] health and safety research,[105] interagency collaboration,[106] and public engagement."[107]

An observation made in the context of health care resonates more broadly about the potential for improved government efforts due to improved information and analysis:

> Big data's transformative potential arises from the information it could generate for many different types of users, including . . . regulators. . . . [Stakeholders in areas such as health care] make countless decisions every day . . . . Those decisions will nearly always turn on the information available to the decision maker. What types of information exist, who is generating that information, and how that information is gathered can have a profound effect on the choices that are made.[108]

The potential value of technological innovations extends beyond informing government decision makers. For example, the information produced by these innovations can promote collaboration among regulated entities and regulators and spur civil engagement by educating interested communities about environmental risks and efforts to reduce them. In these ways, the

102. *See* Mattioli, *supra* note 42, at 543; *see also* SEIZING OPPORTUNITIES, *supra* note 37, at 61 ("Big data technologies are driving enormous innovation while raising novel privacy implications."); Byers, *supra* note 101, at 758 ("[B]ig data enables experimentation, often involving rigorous statistics analyses to identify what option is better.").

103. *See, e.g.*, Breggin & Amsalem, *supra* note 28, at 10,988 (referring to use of data supplied by WaterWatchers to prioritize improvements to city water infrastructure).

104. *See, e.g.*, *id.* at 10,989 (referring to use of meteorological and air-quality data in real time online supplied by the Village Green Project to advance air-quality monitoring).

105. *See, e.g.*, *id.* (referring to information supplied by the California Seafloor Mapping Program to improve maritime safety).

106. *See, e.g.*, *id.* (referring to use of satellite and ground-based observations generated by the Global Earth Observation System of Systems administered by EPA, the National Aeronautics and Space Administration, and the National Oceanic and Atmospheric Administration to coordinate emergency responses to natural and man-made disasters); *id.* at 10,987 (referring to use of geospatial data in the National Wetlands Inventory to integrate maps and supporting data for federal, state, regional, tribal, and local governments, as well as educators and researchers).

107. *Id.* at 10,989, 10,991 (referring to interagency task force efforts to provide utility users access to their own energy data).

108. Madison, *supra* note 80, at 1606; *see also* William G. LeFurgy, *Stewarding Big Data: Perspectives on Public Access to Federally Funded Scientific Research Data, in Big Data – Enabling Big Protection for the Environment* (quoting MANYIKA ET AL., *supra* note 39, at 1–2 (highlighting the promise of big data by pointing to "strong evidence that big data can play a significant economic role to the benefit not only of private commerce but also of national economies and their citizens. Our research finds that data can create significant value for the world economy, enhancing the productivity and competitiveness of companies and the public sector.")), *in* Jayasuriya & Ritscheske, *supra* note 37, at 3.

information supplied by the new technologies can benefit "audiences inside and outside the policy arena."[109]

The need for improved information flows in the context of environmental regulation is particularly acute. Professor Gregg Macey suggests that those who enacted our foundational environmental law infrastructure were "data-starved."[110] The hope is that "big data" and other technology-related innovations will increasingly lead to more informed evidence-based policy making, and also to a more informed body politic.[111]

This Part explores some of the ways in which EPA has begun to put technological innovations to work to improve its compliance and enforcement program. It focuses on EPA's Next Gen initiative, which the agency initiated in 2013 and has touted as a transformative endeavor.[112] As conceptualized by EPA,[113] Next Gen is comprised of five interrelated elements: regulation and permit design, advanced monitoring, electronic reporting, transparency, and innovative enforcement.[114] EPA has recognized the value of newly available information technology in fostering better environmental compliance and supporting enforcement actions in the face of noncompliance. The initiative is designed to involve federal and state regulators, regulated entities, and affected communities in the generation and use of the data these technologies are capable of providing. Subparts A, B, and C, respectively, highlight Next Gen's use of technological innovations by federal and state regulators, regulated entities, and civil society. Subpart D explores the potential of those innovations to increase the transparency of the activities of all stakeholders in the environmental regulatory process.

---

109. Sarah Williams, *More Than Data: Working with Big Data for Civics*, 11 I/S: J.L. & POL'Y FOR INFO. SOC'Y 1, 1 (2015).

110. *See* Macey, *supra* note 63, at 1630–31.

111. Carole Roan Gresenz, *Using Big Data to Assess Community Health & Inform Local Health Care Policymaking* (discussing the potential, and the importance, of evidence-based policy making in the health care policy sphere, and noting that "[t]he gap between the need of local policymakers and non-profit hospitals to . . . understand the health of a population for a refined geographic area and the data available for analysis is often wide; bridging the gap as completely as possible is a central challenge."), *in* Jayasuriya & Ritscheske, *supra* note 37, at 80.

112. EPA announced the new initiative in an article by EPA's Assistant Administrator for Enforcement, Cynthia Giles. *See generally* Giles, *supra* note 19. For an early review of Next Gen, see David L. Markell & Robert L. Glicksman, *Next Generation Compliance*, 30 NAT. RESOURCES & ENV'T 1 (2016).

113. For an alternative model for conceptualizing the design of a regulatory compliance and enforcement program, see *Dynamic Governance, Part I*, *supra* note 29.

114. *See Next Generation Compliance*, EPA, http://www2.epa.gov/compliance/next-generation-compliance (last updated Dec. 23, 2016). EPA has begun issuing regulations and permits that require regulated sources to use Next Gen compliance tools, such as advanced monitoring, electronic reporting, and posting of data on websites available to the public. For a discussion of the agency's efforts to use regulations and permits to advance Next Gen goals, see Hindin & Silberman, *supra* note 45.

A. The Use of Information Technology by Regulatory Agencies

The federal government is a significant generator of data.[115] It also invests significant resources in processing and analyzing this data, although perhaps not as much as agencies need to avoid some of the problems discussed in Part I.[116] The government is taking advantage of modern information technologies in many contexts for many purposes. Satellite technology is being used to track and help analyze a variety of conditions. The National Oceanic and Atmospheric Administration is using data generated by satellite technology to predict the future course of climate change.[117] The federal government uses a geographic information system platform to help develop management policies for lands and resources administered by the Bureau of Land Management and the U.S. Forest Service.[118] As noted above, the National Oceanic and Atmospheric Administration is also integrating geospatial data in geographic information system maps to serve as a resource for environmental emergency responders charged with dealing with events such as oil spills and natural disasters.[119]

EPA is engaged in similar efforts. For example, the agency has compiled an inventory of federal data on power plants in an eGrid that will allow consumers to assess the environmental performance of electricity generators and help them choose the source of their electricity.[120] It has also created a mapping tool to assist in the identification of low-income and minority populations being subjected to disproportionate environmental burdens.[121]

Of direct relevance to environmental compliance and enforcement, EPA is relying on advanced technology in both the monitoring and reporting realms. One technology capable of helping to identify regulatory violations is the infrared camera. EPA has tested a computer program that relies on infrared

---

115. Madison, *supra* note 80, at 1611. The federal government also engages in the collection, aggregation, facilitation, and funding of data generation by nongovernmental sources. *Id.* at 1612–20.

116. *See, e.g.*, Frank Pasquale & Tara Adams Ragone, *Protecting Health Privacy in an Era of Big Data Processing and Cloud Computing*, 17 STAN. TECH. L. REV. 595, 652 (2014) (referring to agency complaints about "the impossible task Congress had set" for collecting and analyzing new data streams in light of resources allocated).

117. *See* Laurie J. Schmidt, *Twelve Years of Satellite Data Help Decode Climate Change*, GLOBAL CLIMATE CHANGE: VITALS SIGNS OF THE PLANET (Apr. 14, 2015), http://climate.nasa.gov/news/2264/twelve-years-of-satellite-data-help-decode-climate-change/; Nat'l Aeronautics and Space Admin., *Taking a Global Perspective on Earth's Climate*, GLOBAL CLIMATE CHANGE: VITAL SIGNS OF THE PLANET, http://climate.nasa.gov/nasa_role/ (last visited Jan. 8, 2017) (stating that "nearly 30 years of satellite-based solar and atmospheric temperature data . . . helped the Intergovernmental Panel on Climate Change" conclude in 2007 that increasing global average temperatures since the mid-twentieth century are very likely due to increased atmospheric concentrations of greenhouse gases, and that National Aeronautics and Space Administration scientists and engineers intend to use data to answer questions such as the future course of temperatures and sea level rise).

118. Breggin & Callan, *supra* note 37, at 131.

119. *See supra* notes 51–52 and accompanying text; Breggin & Callan, *supra* note 37, at 136–37.

120. Breggin & Callan, *supra* note 37, at 132.

121. *Id.* at 132–33.

pollution detection devices to measure emission rates that it expects will be useful in its enforcement efforts.[122] Infrared cameras allow users to detect the presence of compounds that are not visible to the naked eye. The agency itself is already using these cameras to identify methane leaks from oil and gas wells and tanks.[123] Similarly, EPA's New England Regional Laboratory employs solar-powered water quality sensors to measure a variety of pollutant parameters with the aim of identifying the need for further monitoring or targeting sources for enforcement action.[124]

Another prong of EPA's effort to use advances in information technology to foster better compliance relates to its Enforcement and Compliance History Online (ECHO) database. EPA established ECHO in 2002 to help communities assess environmental compliance. In recent years, EPA has modernized ECHO to better support frequent data updates and web services, enhance its interactive features, and improve display on mobile devices.[125] EPA enforcement officials characterize ECHO as a "potential resource to investors and communities,"[126] and claim that it fosters better transparency. Yet EPA has acknowledged that ECHO has not approached its potential value due to persistent data problems.[127] Some commentators who are sympathetic to the ECHO initiative have also characterized ECHO as a tool that has not achieved its potential in increasing transparency.[128] More recently, the Government Accountability Office reported EPA's assertion that the public is making increasing use of the

122. Renee Schoof, *EPA Testing New Way to Measure Air Pollution Emissions*, 46 ENV'T REP. (BNA) 3244 (2015) [hereinafter Schoof, *EPA Testing*]; *see also* Renee Schoof, *Infrared Camera Use Growing in Oil and Gas Sector*, 47 ENV'T REP. (BNA) 1007 (2016) (reporting that Colorado has required the oil and gas industry to detect and reduce methane emissions and has approved the use of infrared cameras to satisfy regulatory monitoring requirements) [hereinafter Schoof, *Infrared Camera*]. EPA enforcement officials have explained that:

> infrared cameras allow users to see dark plumes that look like smoke when volatile organic compounds such as benzene are released to the air even though these emissions are invisible to the naked eye. The EPA uses such cameras to identify methane and other compounds leaking from oil and gas wells, tanks, and other facilities.

Hindin & Silberman, *supra* note 45, at 112.

123. Hindin & Silberman, *supra* note 45, at 112.

124. *Id.* at 112–13.

125. *What's New*, EPA, https://echo.epa.gov/resources/general-info/whats-new (last visited Jan. 8, 2017). For a history of ECHO's evolution and a summary of its benefits and risks, see Lynn L. Bergeson, *ECHO: Enforcement Online, Up Close, and Real Personal,* 12 ENVTL. QUALITY MGMT. 81, 81–83 (2003).

126. Hindin & Silberman, *supra* note 45, at 122.

127. *Known Data Problems*, EPA, https://echo.epa.gov/resources/echo-data/known-data-problems (last visited Jan. 8, 2016) (noting that "EPA has identified some broad-scale data issues that may impact the completeness, timeliness, or accuracy of data shown in ECHO"); *see also Maine Information Relating to US EPA ECHO*, ME. DEP'T OF ENVTL. PROTECTION, http://www.maine.gov/dep/enforcement/echo.html (last visited Jan. 8, 2017) (expressing concerns about ECHO's completeness and accuracy because of differences in vocabulary used by EPA and states, among other factors) [hereinafter Maine ECHO].

128. *See, e.g.*, Clifford Rechtschaffen, *Enforcing the Clean Water Act in the Twenty-First Century: Harnessing the Power of the Public Spotlight*, 55 ALA. L. REV. 775, 803–04 (2004). To be fair, Professor Rechtschaffen offered this critique shortly after ECHO's initial creation.

agency's ECHO website.[129] EPA's own Office of Inspector General concluded in 2016, however, that information obtained from the ECHO website pertaining to the regulation of stationary sources regulated under the CAA was inaccurate, hindering EPA's oversight of delegated state programs and creating a risk of misinforming the public.[130] According to the Inspector General, although twelve million ECHO queries occurred between 2003 and 2012, "[i]naccurate data hinder these activities by misinforming the public about the status of facilities and the level of conducted oversight."[131] Perhaps in response to these criticisms, EPA has taken steps—such as optimizing display of information on mobile devices—to modernize and upgrade ECHO.[132] These and other improvements to the quality and completeness of data in ECHO, achieved through e-reporting and other initiatives, are likely to enhance the value of the database over time.[133]

Another way that EPA intends to leverage data supplied by new information technology relates to its commitment to use data sources outside the agency to support its own enforcement actions. EPA has indicated that it engages in follow-up investigatory and enforcement activities if data supplied by others (such as regulated entities and community groups, as described in the following subparts) raise concerns about potential noncompliance.[134] EPA is particularly likely to do so if it deems the data supplied by those sources to be insufficient in verifying compliance status or supporting enforcement action. That type of follow-up investigation is capable of addressing some of the reliability problems that may accompany the accumulation of data from new information technology, especially if it is produced by nongovernmental sources.

Finally, as explained in the next subpart, EPA has issued regulations requiring electronic reporting by regulated entities under statutes that include the CWA and the Resource Conservation and Recovery Act. In doing so, EPA has created an improved infrastructure to allow regulated entities to report the data to the EPA and states. As the agency has explained, "e-reporting is not just

---

129. U.S. GOV'T ACCOUNTABILITY OFF., GAO-13-115, EPA SHOULD DEVELOP A STRATEGIC PLAN FOR ITS NEW COMPLIANCE INITIATIVE 8 (2012).

130. EPA, REP. NO. 16-P-0164, CLEAN AIR ACT FACILITY EVALUATIONS ARE CONDUCTED, BUT INACCURATE DATA HINDER EPA OVERSIGHT AND PUBLIC AWARENESS, AT A GLANCE 9 (2016), https://www.epa.gov/sites/production/files/2016-05/documents/20160503-16-p-0164.pdf. The Inspector General found, for example, that of sixty-five facilities listed as major operating facilities, 26 percent were either closed, minor sources, never constructed, or not a facility. *Id.* EPA's response to the OIG report is included as Appendix B.

131. *Id.* at 12.

132. *See* EPA, *supra* note 125.

133. Markell & Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming), *supra* note 29.

134. *Cf.* Hindin & Silberman, *supra* note 45, at 113 (noting that buoy measurements for water pollution parameters by an EPA regional office "do not directly indicate regulatory noncompliance but they can help support follow-up compliance monitoring or targeting"); *id.* at 115 & n.106 (noting the need for follow-up inspections in connection with third-party verification and certification approaches).

converting paper to an electronic media. It is rather a system that guides the user through the reporting process with integrated compliance assistance and data quality checks."[135] EPA's 2015 National Pollutant Discharge Elimination System (NPDES) e-reporting rule has the potential to influence EPA enforcement priorities. Among other things, the rule's coverage of non-major facilities will give EPA access to information about the compliance status of many regulated parties for the first time.[136]

In sum, EPA is in the process of creating programs and infrastructure that will enable it to take advantage of advanced information technology to create a more effective set of compliance and enforcement mechanisms. Initiatives such as the e-reporting rules make it clear that the burden of generating the data made available by these technologies will not fall entirely on EPA or its state counterparts. Regulated entities will also have a significant role to play. The next subpart addresses that role.

B. The Use of Information Technology by Regulated Entities

1. Monitoring

Monitoring regulated facilities to ascertain compliance status presents logistical problems. In the air pollution context, one traditional approach was to conduct stack tests to determine if emissions were consistent with applicable permit limits.[137] Stack tests, however, are recognized to be a less-than-ideal tool to assess compliance. A recent decision by the Court of Appeals for the D.C. Circuit describes the limits of stack testing to establish regulatory emission limits for hazardous air pollutants under the CAA and determine compliance with those limits:

---

135. EPA, *Priority Next Generation Compliance Research Questions* 9 (May 18, 2016) (on file with authors).

136. *See, e.g.*, National Pollutant Discharge Elimination System (NPDES) Electronic Reporting Rule, 80 Fed. Reg. 64,064, 64,064 (Oct. 22, 2015):

> [T]he final rule requires authorized NPDES programs to share the minimum set of NPDES program data (appendix A to 40 CFR part 127) with EPA for all facilities including non major facilities. Historically, EPA and authorized NPDES programs have focused on major facilities as a way of prioritizing resources for permitting, enforcement and data sharing. Over time, there has been a growing recognition that these nonmajor sources significantly impact water quality as well.

EPA has issued guidance on the recipients of the information to be generated by electronic NPDES reporting. *See* NPDES Electronic Reporting Rule Implementation Guidance, 81 Fed. Reg. 62,395 (Sept. 9, 2016) (to be codified at 40 C.F.R. pts. 9, 122, 123, 124, 127, 403, 501 & 503). For more thorough analysis of this potentially very significant development, see Markell & Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming), *supra* note 29.

137. *See* F. William Brownell, *"Regulation by Guidance": A Response to EPA*, 10 NAT. RESOURCES & ENV'T 56, 57 (1996) (stating that in the early 1970s, EPA relied on stack tests "[b]ecause continuous emission monitors (CEMs) were unavailable").

> Further complicating the task [of regulating hazardous air pollutant emissions] is the way in which sources typically measure emissions. Virtually all of the data the EPA collects to set [technology-based regulatory] floors come from the three-run stack test. The three-run stack test, as the name suggests, involves three measurements of the source's emissions taken over a short time period (i.e., no more than a few days) with each of the three test "runs" lasting from one hour to four hours. Because the tests provide three "snapshots" of a source's emissions performance, they cannot accurately represent the source's full range of emissions over all times and under all conditions.[138]

A lawyer with an environmental group characterized stack tests in a very critical way:

> Environmental enforcement relies almost entirely on industry's own monitoring, but too much of that monitoring—especially under the CAA—is a sham. Compliance with hourly emission limits for some pollutants is tested every other year—and sometimes less often—through three-hour stack tests that are too easy to manipulate to obtain favorable results.[139]

One court, in a CAA enforcement action, noted that "there is little doubt that had stack tests been performed with greater regularity . . . a substantial number of additional violations might have been identified."[140]

EPA has taken several actions to address concerns about stack tests. In addition to refining the protocol for evaluating stack tests in some circumstances,[141] it has required the use of continuous emissions monitoring (CEM) in others to assess compliance with regulatory standards.[142] CEM has the potential to provide a more accurate depiction of compliance status over time. EPA enforcement officials have described CEM, which usually is used to monitor compliance by stationary sources with air pollution controls,[143] as

138. U.S. Sugar Corp. v. EPA, 830 F.3d 579, 632 (D.C. Cir. 2016) (emphasis omitted) (citations omitted); *see, e.g.*, Paul D. Hoburg, *Use of Credible Evidence to Prove Clean Air Act Violations*, 25 B.C. ENVTL. AFF. L. REV. 771, 815 (1998) (noting that stack tests "covered brief periods of time and yielded short-term 'snapshots' of the source's emissions"); EPA, *supra* note 135, at 3 ("EPA and states write permits allowing facilities to emit or discharge certain levels of pollutants into the air or water. Companies are typically required to monitor levels of pollution to ensure that they are under their permit limits. This type of monitoring is generally on a periodic basis, such as a daily grab sample, monthly averages based on weekly grab samples, or just once a month or even annually or less.").

139. Eric Schaeffer, *A Fresh Start for EPA Enforcement*, 38 ENVTL. L. REP. 10,385, 10,387 (2008); *see also* James Miskiewicz & John S. Rudd, *Civil and Criminal Enforcement of the Clean Air Act After the 1990 Amendments*, 9 PACE ENVTL. L. REV. 281, 361 (1992). For discussion of the inadequacies of traditional stationary monitoring technologies, see McGarity, *supra* note 68, at 1478–79.

140. United States v. SVM Corp., 667 F. Supp. 1110, 1126 (D. Md. 1987).

141. *U.S. Sugar Corp.*, 830 F.3d at 635–37 (noting that EPA decided to use an "upper prediction limit", whose validity the court upheld, to establish regulatory standards, and that the upper prediction limit "produces a range of values that is expected, given the variance in the relevant stack-test data, to encompass the average emissions levels achieved by the best performing sources a specified percentage of the time") (emphasis omitted).

142. *Id.* at 654.

143. According to EPA, "[a] continuous emission monitoring system (CEMS) is the total equipment necessary for the determination of a gas or particulate matter concentration or emission rate

monitoring that "measures emissions sufficiently frequently to provide a representative measure of the monitored unit's continuous emission levels under the applicable rules."[144] The agency successfully relied on CEM to track emissions of sulfur dioxide and oxides of nitrogen by power plants subject to the CAA's acid rain control requirements.[145]

Based at least partly on this experience, EPA has stressed the value of shifting from periodic to continuous emissions monitoring and reporting on a more widespread basis to provide better data that represent actual conditions and to identify violations more quickly. EPA's increasing resort to CEM derives from its conviction that "[b]y promoting high regulatory compliance, the use of CEMS contributed to increased certainty for industry with significantly less regulator and industry time spent on enforcement cases."[146] EPA has identified CEM as a tool for promoting high compliance levels, opining that it "may be feasible for use in a broader range of regulatory settings" as its cost falls.[147] Others also see the potential for CEM to provide greater reliability and greater credibility.[148] Moreover, as new technology develops, CEM systems become increasingly available "for a broader range of air emissions, including toxic substances, and water pollutants."[149] As CEM technology advances, so, too, will its reliability and value as a tool to enhance compliance and enforcement.

Another way to make monitoring a more useful tool for identifying and addressing noncompliance is to expand the locations where it is done. EPA has

---

using pollutant analyzer measurements and a conversion equation, graph, or computer program to produce results in units of the applicable emission limitation or standard." *Air Emission Measurement Center (EMC): Continuous Emission Monitoring Systems*, EPA, https://www.epa.gov/emc/emc-continuous-emission-monitoring-systems (last updated Sep. 27, 2016).

144. Hindin & Silberman, *supra* note 45, at 116; *see also* EPA, *supra* note 135, at 3 ("CEMs measure emissions sufficiently frequently to provide a representative measure of the monitored unit's continuous emission levels under the applicable rules."). CEM "generally takes one of two forms: (1) a continuous parameter monitor, which measures, e.g., a source's temperature, pressure or oxygen content; or (2) a continuous emissions monitor, which measures the pollutant concentration in the source's emissions." *U.S. Sugar Corp.*, 830 F.3d at 654.

145. *See* John Schakenbach et al., *Fundamentals of Successful Monitoring, Reporting, and Verification under a Cap-and-Trade Program*, 56 J. OF THE AIR & WASTE MGMT. ASS'N 1576, 1576–77 (2006) (discussing the use of continuous monitoring and approaches to produce a successful monitoring regime); Lesley K. McAllister, *Enforcing Cap-and-Trade: A Tale of Two Programs*, 2 SAN DIEGO J. CLIMATE & ENERGY L. 1, 4–7 (2010) (describing how CEM equipment and automatic verification systems bolstered compliance levels under the CAA's acid rain program); Hindin & Silberman, *supra* note 45, at 116 (concluding that the use of CEM under the acid rain control program "proved instrumental in ensuring that the Program's mandated reductions . . . were achieved").

146. EPA, *supra* note 135, at 3. EPA nevertheless has expressed its interest in further research on "whether and how including real-time and/or continuous monitoring in permits impacts the behavior of the regulated facilities." *Id.*

147. Hindin & Silberman, *supra* note 45, at 116.

148. *See, e.g.*, Wendy E. Wagner, *Commons Ignorance: The Failure of Environmental Law to Produce Needed Information on Health and the Environment*, 53 DUKE L.J. 1619, 1691, 1691 n.250 (2004); Arnold W. Reitze, Jr. & Steven D. Schell, *Self-Monitoring and Self-Reporting of Routine Air Pollution Releases*, 24 COLUM. J. ENVTL. L. 63, 116 (1999).

149. Hindin & Silberman, *supra* note 45, at 116.

begun requiring regulated entities to monitor conditions in locations that previously were not routinely monitored. One prominent example is fenceline monitoring, which EPA officials describe as "the strategic placement of monitoring equipment at locations along or adjacent to facility property lines to detect, identify, and quantify pollutant releases from point sources and fugitive emissions at regulated facilities."[150] In a "Draft Roadmap for Next-Generation Air Monitoring" published in 2013, EPA identified three goals for the use of advanced monitoring techniques in different locations, including fenceline monitoring:

- Promote development of affordable, near source, fenceline monitoring technologies and sensor network-based leak detection systems . . . .
- Supplement air quality monitoring networks through development of low cost, reliable air quality monitoring technology . . . .
- Support environmental justice communities and citizen efforts to measure air pollution in local areas.[151]

EPA has taken steps to implement this agenda. Recent EPA enforcement actions, settlements, and regulations have included requirements that regulated parties monitor at their facility fencelines.[152] For example, in 2015, EPA issued final regulations that require petroleum refineries to deploy passive fenceline monitoring to measure facility emissions of benzene.[153] The regulations also specify procedures for subtracting background concentrations and contributions to fenceline concentrations from other sources and mandate corrective action if an applicable fenceline benzene concentration action level is exceeded.[154]

150. *Id.* EPA has noted the development of a new technology, differential absorption light detection, which can produce more accurate measurements of fugitive emissions from tanks. EPA, *supra* note 135, at 2. For further discussion of fenceline monitoring, see *supra* note 68 and accompanying text.

151. EPA, DRAFT ROADMAP FOR NEXT-GENERATION AIR MONITORING 2 (2013), http://www.eunetair.it/cost/newsroom/03-US-EPA_Roadmap_NGAM-March2013.pdf. The water pollution analog of CEM is whole effluent toxicity, which assists regulators in identifying discharges of toxic pollutants that threaten the ability of aquatic organisms to survive and reproduce. Hindin & Silberman, *supra* note 45, at 117.

152. For a recent example of an enforcement action that includes advanced monitoring in new locations, see *Tesoro and Par Clean Air Act Settlement*, EPA, https://www.epa.gov/enforcement/tesoro-and-par-clean-air-act-settlement (last updated July 18, 2016) (describing a 2016 settlement that committed the defendants to use infrared gas-imaging cameras at four refineries to supplement the company's enhanced leak detection and repair program). EPA noted that "[t]hese cameras are able to locate fugitive VOC emissions that may not be otherwise detected." *News Releases from Region 10: Oil Refiners to Reduce Air Pollution at Six Refineries Under Settlement with EPA and Department of Justice*, EPA (July 18, 2016), https://www.epa.gov/newsreleases/oil-refiners-reduce-air-pollution-six-refineries-under-settlement-epa-and-department; *see also* Delmar R. Ehrich, *Preparing for the Imminent Rise of Citizen Science*, LAW360, (June 24, 2016, 12:21 PM), http://www.faegrebd.com/files/121255_Preparing_For_The_Imminent_Rise_Of_Citizen_Science.pdf. We review the expanding use of fenceline monitoring in more detail in Markell & Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming), *supra* note 29.

153. Refinery Risk, *supra* note 100, at 75, 191–200, 75, 254–57.

154. *Id.* at 75,192; *see also* Hindin & Silberman, *supra* note 45, at 116 (noting that fenceline monitoring can "serve as [a] trigger[] for further monitoring or corrective actions by the facilities").

EPA officials explain the agency's rationale for its recent issuance of regulations requiring fenceline monitoring as follows:

> Environmental monitoring traditionally occurs within facility fencelines where the physical locations of the monitors correspond to stacks, sources, units, and equipment subject to standards or limits. Today, however, concerns have increased regarding impacts regulated facilities may have on surrounding communities and public health due to excess emissions, undetected releases (planned or unexpected), or noncompliance, generally, with all of a facility's regulatory requirements. Due to these concerns, regulators and sources are increasingly employing fenceline, remote, and ambient monitoring alongside, adjacent to, or further outside facility property lines.[155]

As the foregoing discussion indicates, features of traditionally available pollution monitoring technologies limit the value of the data they produce. Their measurements tend to be sporadic and fixed at the emission point, creating significant constraints on the government's, regulated parties', and the community's capacity to measure regulated entities' performance and accurately detect regulatory violations. New technologies, including CEM and devices that allow accurate fenceline monitoring, have the potential to reduce those governance challenges. These developments are likely to impact key stakeholders of all kinds. As agency enforcement officials put it recently, "regulators must use and promote advanced pollutant detection technology so regulated entities, the government, and the public can more easily 'see' pollutant discharges, environmental conditions, and noncompliance."[156] The result should be greater accountability for everyone, provided that regulators thoughtfully manage and analyze the data generated by the new technologies.

2. *Reporting*

EPA is also requiring regulated entities to use advanced technologies to report on environmental performance. Next Gen's electronic reporting (or e-reporting) component involves a shift by EPA "away from outdated paper reporting toward e-reporting" by regulated entities.[157] The agency has high hopes for e-reporting, conceiving of it as a way to facilitate compliance and track reporting in many ways. EPA believes that a shift to e-reporting will

---

Regulations adopted under the CWA require whole effluent toxicity testing under certain conditions. 40 C.F.R. § 122.21(j)(5) (2015); 40 C.F.R. § 122.44(d)(1)(iv) (2015).

155. Hindin & Silberman, *supra* note 45, at 116.

156. *Id.* at 106. For example, "high-quality monitoring data can be used to trigger corrective action where predictive data show a performance trend above a regulated unit's usual or preferred performance level . . . ." *Id.* at 113. Some believe that drones have potential value in environmental enforcement "because they can be packed in a suitcase and deployed quickly," but EPA does not currently plan to use them for enforcement purposes due to legal barriers and liability issues. Renee Schoof, *Drone Use for Environmental Monitoring May Grow under Rule*, 47 DAILY ENV'T REP. (BNA) 2609 (2016) (quoting a remote sensing specialist).

157. Hindin & Silberman, *supra* note 45, at 118.

minimize errors introduced through manual data entry, prompt the development by the private sector of e-reporting technology that is easier and cheaper to use, facilitate "electronic data checks" that allow self-correction by regulated entities and flag inconsistent or impossible entries, and help government provide compliance assistance to regulated entities.[158] According to EPA, "e-reporting can reduce the costs associated with paper reporting and allow regulated entities, government agencies and the public to more quickly identify and address violations."[159] It can also promote transparency, another key Next Gen component.[160] According to an enforcement official intimately involved in the development and roll-out of Next Gen, "[g]reater accessibility could also drive better compliance performance as facilities learn from each other about what performance is possible."[161] It can also enable data mining on the reports.

To take advantage of these benefits of e-reporting, EPA has already issued regulations requiring electronic reporting under the CWA's NPDES permit program.[162] Pursuant to legislation adopted in 2012 to amend the Resource Conservation and Recovery Act,[163] it also has established a national electronic manifest system to better track the location and condition of hazardous waste from the point at which it leaves the generating site to its ultimate disposal.[164] E-reporting can be fully effective, however, only if the software tools provided for its use properly guide regulated entities in submitting accurate and complete data, and if the information supplied accurately reflects regulated party performance and any government response.[165] EPA has stated that its further study of the accuracy of self-reported data,[166] and experience with existing regulatory programs, has demonstrated that reasons exist to be wary of assuming that all self-reported data is accurate.[167] Differences in EPA and state

158. Giles, *supra* note 19, at 25.

159. EPA, *supra* note 135, at 9.

160. *See* Hindin & Silberman, *supra* note 45, at 118 ("E-reporting promotes compliance by giving regulators—and through regulators, the public—timely access to high quality, complete, and consistent compliance information.").

161. Giles, *supra* note 19, at 25; *see also* EPA, *supra* note 135, at 9 (asserting that the transparency that results from e-reporting "could drive compliance by making relevant information easily accessible to regulators and the public").

162. National Pollutant Discharge Elimination System (NPDES) Electronic Reporting Rule, 80 Fed. Reg. 64,064 (Oct. 22, 2015) (to be codified at 40 C.F.R. pt. 127).

163. The Hazardous Waste Electronic Manifest Establishment Act, Pub. L. No. 112-195, 126 Stat. 1452 (2012).

164. *Hazardous Waste Electronic Manifest System (E-Manifest)*, EPA, https://www.epa.gov/hwgenerators/hazardous-waste-electronic-manifest-system-e-manifest/ (last updated Jan. 6, 2016); Hazardous Waste Management System; User Fees for the Electronic Hazardous Waste Manifest System and Amendments to Manifest Regulations, 79 Fed. Reg. 7518-01 (Feb. 7, 2014) (to be codified at 40 C.F.R. pts 262, 263, 264, 265 & 271).

165. *See* Hindin & Silberman, *supra* note 45, at 118.

166. EPA, *supra* note 135, at 7 (indicating, however, that studies on self-reported wastewater discharge monitoring data "generally, do not indicate a likelihood of widespread inaccurate or fraudulent monitoring self-reporting").

167. *See, e.g.*, Robert L. Glicksman, *Regulatory Safeguards for Accountable Ecosystem Service Markets in Wetlands Development*, 62 U. KAN. L. REV. 943, 952–53 (2014) (describing fraudulent

vocabulary, among other factors, suggest the need for additional work to improve accuracy with respect to both government actions and regulated party performance.[168]

EPA is not alone in moving toward greater reliance on e-reporting. Ohio's environmental agency has adopted a mandatory electronic reporting system for discharge monitoring reports under the CWA. An EPA report suggests that this innovative effort has had considerable success. The system generated a 99 percent reporting rate, which resulted in a 90 percent decline in reporting errors.[169] It also allowed the agency to reallocate staff members away from reporting oversight responsibilities to other areas of need.[170] EPA officials concluded that the Ohio effort demonstrated the potential of electronic reporting to improve the accuracy of the information the state agency uses in making compliance- and enforcement-related decisions and enabled the agency to administer its compliance program more efficiently.[171]

Immediate feedback technology is another technology that EPA regards as a promising compliance enhancement tool. As EPA enforcement officials have described it, this technology

> provid[es] regulated entities with accurate measures, in a standardized format, of deviations indicating that regulatory requirements are being, or may soon be, violated.... Regulated entities can receive real-time performance feedback and data intended to prompt, automatically or through user responses to the alerts, remedial actions to correct or prevent violations.[172]

These officials hope that these mechanisms will yield "positive behavioral impacts."[173] Among other impacts, immediate feedback technology is likely to create in regulated entities a perception of increased risk of detection of

---

misreporting of renewable fuel credits under a regulatory program implemented under the Energy Policy Act of 2005); *see also* Giles, *supra* note 19, at 26 ("And where government relies on self-reporting for compliance data, we also need ways to check for accuracy.").

168. *See, e.g.*, EPA, *supra* note 127 (expressing concerns about ECHO's completeness and accuracy because of differences in vocabulary used by EPA and states, among other factors).

169. Hindin & Silberman, *supra* note 45, at 118 n.137.

170. *Id.*

171. *Id.* at 118.

172. *Id.* at 111. Based on a study of the use of mobile-monitoring technologies by Texas air quality regulators to track hazardous air pollutant emissions, Professor McGarity concluded nearly a decade ago that:

> the key to effective mobile monitoring is the advent of sophisticated ambient-air-quality sampling devices that are capable of providing 'real-time' measurements of ambient concentrations of multiple [hazardous air pollutant]s. Unlike traditional stationary monitoring devices, in which samples are collected over a period of time and sent to laboratories for subsequent analysis, these modern devices provide immediate feedback to the team members.

McGarity, *supra* note 68, at 1457. Such feedback may be able to help regulated entities manage their operations to increase the chances of meeting compliance targets.

173. Hindin & Silberman, *supra* note 45, at 111.

noncompliance, a "classic . . . deterrence response."[174] The technology has been used successfully in non-environmental contexts, such as traffic control. Studies have found that speed boards that electronically display a driver's speed in relation to the speed limit induce motorists to drive slowly and more carefully.[175] In the environmental enforcement context, EPA has relied on a version of immediate feedback technology in a settlement that requires the operator of a petroleum company to install electronic release detection monitoring equipment for underground storage tanks at its gas stations. The data supplied by this equipment will provide around-the-clock surveillance from a central location of environmentally dangerous leaks.[176] The potential obstacles to use of the technology as a compliance promoter include unresolved questions concerning its cost-effectiveness, the accuracy and reliability of its results, and EPA's legal authority to require its use.[177] If those questions yield positive answers, EPA is likely to increase its use of such innovations through regulations or enforcement settlements.[178]

C. *The Use of Information Technology by Civil Society*

1. *The Rise of Citizen Science*

Increased use of advanced technologies is not limited to environmental agencies or regulated entities.[179] As the discussion below indicates, individuals, citizen groups, and communities increasingly have access to affordable sensors and other monitoring devices capable of generating data on both ambient environmental conditions and regulatory compliance.[180] According to researchers who have extensively surveyed the use of such devices in the environmental context, "[c]ollecting information from the general public . . . is resulting in large amounts of data generated through apps and websites that enable the public to contribute to growing stores of environmental data."[181]

---

174. *Id.* at 105.

175. *Id.* at 109 n.49.

176. *Id.* at 111.

177. *Id.*

178. *See id.* EPA has already required the installation of advanced electronic release detection monitoring equipment at gas stations with underground storage tanks in a settlement reached with Total Petroleum Puerto Rico Corporation in 2015. *See id.*

179. Terry, *supra* note 39, at 389–90. For a detailed chart on the databases, tools, and initiatives that qualify as the use of big data in the environmental context by all levels of government and the private sector, see Breggin & Amsalem, *supra* note 28, at 10,987–91. *See also* ELI, BIG DATA, *supra* note 39, at 3–29 (also including data generation efforts by nongovernmental organizations).

180. *See infra* notes 185–189 and accompanying text.

181. ELI, BIG DATA, *supra* note 39, at 23. Although this Part focuses on the generation of data by citizens and community groups that may enhance compliance promotion and enforcement efforts, citizen groups are also likely to access information generated by industry through tools such as electronic reporting and to use that information to pressure industry to improve performance or as the evidentiary foundation for citizen suits against alleged violators. For further discussion of this use of data made available by advanced information technologies, see *Dynamic Governance, Part I*, *supra* note

Some refer to the phenomenon as "citizen science."[182] The federal government recognizes the value of the participation of individuals, community groups, and others in providing useful information to the government. This recognition is reflected in its creation of an official website, citizenscience.gov. The site's purpose is

> to accelerate the use of crowdsourcing[183] and citizen science across the U.S. government. The site provides a portal to three key assets for federal practitioners [and others]: a searchable catalog of federally supported citizen science projects, a toolkit to assist with designing and maintaining projects, and a gateway to a federal community of practice to share best practices.[184]

29, at Parts III and IV, and Markell & Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming), *supra* note 29, at Parts IIC and III.

182. ELI, CLEARING THE PATH, *supra* note 76, at 5. The Environmental Law Institute defines citizen science as "a form of open collaboration where members of the public undertake scientific work, often in collaboration with professional scientists and scientific institutions, to meet real world goals." *Id.* For a more in-depth review of the role of citizens in environmental enforcement, see *Dynamic Governance, Part I*, *supra* note 29, at 618–29. For discussion of the role of citizen science in public policy making, see generally ELI, CLEARING THE PATH, *supra* note 76.

183. A contributing editor for Wired magazine has defined "crowdsourcing" as "[t]he act of taking a job once performed by a designated agent (an employee, freelancer or a separate firm) and outsourcing it to an undefined, generally large group of people through the form of an open call, which usually takes place over the Internet." William Safire, *Fat Tail*, N.Y. TIMES (Feb. 5, 2009), http://www.nytimes.com/2009/02/08/magazine/08wwln-safire-t.html?_r=3&ref=magazine&.

184. CITIZENSCIENCE.GOV, http://www.citizenscience.gov/ (last visited Jan. 8, 2017) [hereinafter citizenscience.gov]. As of July 8, 2016, the catalog, which is searchable by agency, included no links to EPA programs through that search technique. It did list 116 projects sponsored by sixteen federal and state agencies that related to ecology and environment, however, and one of those was EPA's Urban Waters Program—Amigos Bravos. *Federal Crowdsourcing and Citizen Science Catalog: EPA Urban Waters Program—Amigos Bravos*, citizenscience.gov, https://ccsinventory.wilsoncenter.org/#projectId/132 (last visited Jan. 8, 2017).

A nongovernment initiative in the same vein is eBird, an effort by a global network of volunteers to collect information about the distribution and abundance of birds that "has evolved from a stand-alone citizen-science project focused on collecting data, into a cooperative partnership involving several distinct user groups spanning multiple scientific domains and dozens of partner organizations." Brian L. Sullivan et al., *The eBird Enterprise: An Integrated Approach to Development and Application of Citizen Science*, 169 BIOLOGICAL CONSERVATION 31, 32 (2014), http://www.sciencedirect.com/science/article/pii/S0006320713003820. The design of the project was influenced by studies on how to improve the impact and utility of citizen-science data. These studies recommended:

> incorporating more rigorous sampling techniques into unstructured data collection processes, improving data quality, broadening the data-user community, and improving communication between those using the data and those collecting it. We suggest that achieving these objectives is best accomplished by expanding the range of activities routinely encompassed with the running of citizen-science projects to extend beyond data collection to include community engagement, data curation, data synthesis and analysis, pattern visualization, and delivery of results to a broad community of possible stakeholders.

Id.

Increased access to data-generating technology is one reason that data generated by nongovernmental sources has begun to proliferate.[185] Access has increased because the cost of the devices that generate the data has fallen sharply:

> In this era of big data . . . the technical constraints on computing have loosened, allowing data to be more easily collected, stored, and analyzed. The lower cost associated with these tasks has allowed data to get even bigger and has made data-intensive analyses much more feasible in many settings.[186]

Many different devices can generate data of potential value in the implementation of government programs. These include sensors that "record the position, time, and basic attributes of a mechanical" device performing some function; devices that individuals can activate (such as smart cards); and devices like mobile phones that can link users to different applications.[187] In the environmental arena, for example, EPA operates solar-powered water quality sensors that take measurements every fifteen minutes and upload the results to the agency's public website.[188] The increased access to this array of technology has prompted what some refer to as the "democratization" of the collection of data of potential value to the implementation of regulatory programs.[189] EPA's Next Gen initiative and similar endeavors by state

185. *See* Fahey, *supra* note 39, at 329 (referring to "decades of advances in hardware development and high performance computing" that make the technology more accessible as a "critical step . . . in the process of democratizing big data analysis").

186. Madison, *supra* note 80, at 1610; *see also* Fahey, *supra* note 39, at 330 ("[T]he major news item regarding the democratization of big data is that it is now much cheaper to store and analyze data. Small businesses can now afford the analytical tools, services and experts whereas before this storage and analysis was prohibitively expensive. One of the benefits, then, may be that the U.S. Government, like the small business owner, can store and analyze more data at lower cost. This is magnificent news to those who are wrestling with budgets."); Breggin & Amsalem, *supra* note 28, at 10,985 (citing Emily G. Snyder et al., *The Changing Paradigm of Air Pollution Monitoring*, 47 ENVTL. SCI. & TECH. 11,369 (2013)); Rodriguez, *supra* note 50 (quoting Assistant Administrator Cynthia Giles that as new monitoring technologies "become better, cheaper, smaller, more mobile—all of which is happening today—they're going to be in much, much, wider use in the future"). Air quality monitoring devices can cost as little as $150 to $200. Main, *supra* note 64; *see also* Ehrich, *supra* note 152 ("[A]dvances in information and monitoring technologies increasingly put portable, small, lower-cost monitoring devices into the hands of individuals or groups interested in air or water quality in their personal environments.").

187. Batty, *supra* note 71, at 131. Some sensors are designed to prompt actions by individuals to protect their own health as well as provide data for use in enforcement actions. A group of entrepreneurs and scientists, for example, has tested a wearable device that, when connected to a smart phone app, can provide location-specific measurements of air quality. Sherrell Dorsey, *This Wearable Device Helps You Ditch Air Pollution*, TRIPLE PUNDIT (June 8, 2015), http://www.triplepundit.com/2015/06/wearable-device-helps-ditch-air-pollution/.

188. Hindin & Silberman, *supra* note 45, at 112.

189. Fahey, *supra* note 39, at 329 (discussing Google's contributions to this phenomenon). "An air quality sensor can fit in the palm of your hand. Every ten or twenty seconds, it can detect substances without the need to send samples to a lab. These devices are evolving at a rapid pace." Macey, *supra* note 63, at 1649. Analysis of this phenomenon is not confined to data relating to environmental conditions. "[T]here is a growing literature on the democratization of science, including work by social

environmental regulators reflect efforts to institutionalize the use of citizen science as a means of enhancing compliance and enforcement, as the next subpart illustrates.

2. The Role of Citizen Science in Environmental Enforcement

Both federal and state agencies have welcomed the development and increased availability of these technologies and encouraged citizens to participate in these kinds of efforts. EPA Administrator Gina McCarthy, for example, has touted the use of smart phones as air quality monitors.[190] Community-driven generation of information on environmental conditions already contributes to federal and state government efforts across the nation to implement and enforce the environmental laws. Government officials have encouraged or made use of information generated by various citizen initiatives, including those designed to promote compliance with regulatory obligations. For example, the "Virginia Department of Environmental Quality has established differing levels of required data quality depending upon the intended use of citizen-science monitoring data."[191] Colorado officials have used information supplied by infrared cameras operated by workers to detect methane leaks in natural gas wells.[192] Officials in Boston have used data supplied by citizen-operated monitors to detect fugitive air emissions.[193] University researchers have engaged in monitoring in vans to identify the sources of benzene emissions in Houston neighborhoods.[194] Volunteers organized by the University of Nebraska-Omaha have taken water quality samples along the Mississippi River in an effort to detect and determine the sources of atrazine pollution.[195] An iPhone application that IBM developed makes it possible for citizens to monitor water quality.[196] The Freshwater Trust, an Oregon-based environmental nongovernmental organization, has used

---

scientists and educators studying the best ways to invite and support lay researchers into science." Mary L. Lyndon, *The Environment on the Internet: The Case of the BP Oil Spill*, 3 ELON L. REV. 211, 231 (2012).

190. Pat Rizzuto, *Get Ready for Phone Air Monitor Data, EPA's McCarthy Says*, 47 ENV'T REP. (BNA) 1029 (2016). At least one non-environmental federal agency has actually created a smart phone app to solicit the public's help in identifying violators of the law. *See Operation Predator App*, U.S. IMMIGRATION AND CUSTOMS ENFORCEMENT, https://www.ice.gov/predator/smartphone-app (last visited Jan. 8, 2017) (for use in apprehending child predators).

191. ELI, CLEARING THE PATH, *supra* note 76, at 31.

192. William Yardley, *New Technology Is Keeping the Air We Breathe Under an Unprecedented Level of Scrutiny*, L.A. TIMES (Oct. 18, 2015, 4:00 AM), http://www.latimes.com/nation/la-na-sej-measuring-future-20151018-story.html.

193. *See id.*

194. Among other things, the researchers discovered that leaks from crude oil and natural gas pipelines were contributing more to the problem than had been realized. Dianna Wray, *The Way We Currently Monitor Air Pollution Near the Ship Channel Sucks, Researchers Say*, HOUSTONPRESS (Mar. 28, 2016, 6:00 AM), http://www.houstonpress.com/news/the-way-we-currently-monitor-air-pollution-near-the-ship-channel-sucks-researchers-say-8263528.

195. Rejeski & McElfish, *supra* note 67, at 64.

196. Hindin & Silberman, *supra* note 45, at 113 n.93.

boats equipped with cameras linked to Google Maps to assess surface water quality.[197] Even if these initiatives were not initially designed to investigate and provide evidence to pursue enforcement of regulatory violations, the information they generate could well be useful for those purposes.[198]

Because quality control over citizen science is unlikely to match an environmental agency's own information-gathering efforts, agencies have used data from nongovernmental sources as a signal warranting their own further inquiries into compliance status or ambient conditions. EPA, for example, expanded on-road emissions testing instead of relying exclusively on laboratory tests in the wake of the scandal provoked by Volkswagen's installation of "defeat devices" in its cars that turned emission control equipment on when lab tests were being conducted but off when the cars hit the road.[199] The use of emissions testing equipment by environmental groups and independent laboratories had already increased the chance that these kinds of violative practices would come to light, as they did in the Volkswagen case itself.[200]

In one of the clearest examples of the potential for citizen science to enhance environmental enforcement, air quality sampling by community activists in Tonawanda, New York prompted the state Department of

---

197. Frederick Reimers, *Mapping America's Disgusting Waterways*, BLOOMBERG BUSINESSWEEK (Oct. 15, 2015, 1:03 PM), http://www.bloomberg.com/news/articles/2015-10-15/mapping-america-s-disgusting-waterways. These kinds of initiatives are not confined to either the United States or to pollution control laws. Satellite technologies, for example, are helping to identify industrial fishing activities that harm existing stocks. Global Fishing Watch has developed a product that allows anyone to view and interact with data on fishing across the globe. Douglas McCauley, Opinion, *How Satellites and Big Data Can Help Save the Oceans*, ENVIRONMENT360 (Apr. 13, 2016), http://e360.yale.edu/feature/how_satellites_and_big_data_can_help_to_save_the_oceans/2982/.

> Global Fishing Watch is the product of a technology partnership between SkyTruth, Oceana, and Google that is designed to show all of the trackable fishing activity in the ocean. This interactive web tool—currently in prototype stage—is being built to enable anyone to visualize the global fishing fleet in space and time. . . . Global Fishing Watch will be available to the public, enabling anyone with an internet connection to monitor when and where commercial fishing is happening around the globe.

GLOBAL FISHING WATCH, https://www.greenpolicy360.net/w/Global_Fishing_Watch (last visited Jan. 31, 2017) (emphasis omitted). In addition, with funding by the Bureau of Land Management, volunteers have helped track sage grouse populations on public lands. Rejeski & McElfish, *supra* note 67, at 64; *see also supra* note 184 and accompanying text (discussing the eBird enterprise for collecting data on bird distribution and abundance).

198. For example, if monitoring of ambient conditions in Houston to determine whether neighborhoods are being exposed to unsafe levels of benzene emissions identifies the likely sources contributing to excessive ambient benzene concentrations, that information may spur further investigation by federal or state regulators to determine if those sources are violating permit limits.

199. Danny Hakim & Jad Mouawad, *Galvanized by VW Scandal, E.P.A. Expands On-Road Emissions Testing*, N.Y. TIMES (Nov. 8, 2015), http://www.nytimes.com/2015/11/09/business/energy-environment/epa-expands-on-road-emissions-testing-to-all-diesel-models.html?_r=0.

200. Evidence of irregularities in Volkswagen cars surfaced as a result of testing performed at West Virginia University by a nonprofit group, the International Council on Clean Transportation. Jack Ewing, *VW Presentation in '06 Showed How to Foil Emissions Tests*, N.Y. TIMES (Apr. 26, 2016), http://www.nytimes.com/2016/04/27/business/international/vw-presentation-in-06-showed-how-to-foil-emissions-tests.html.

Environmental Conservation to conduct follow-up studies which detected unsafe concentrations of benzene linked to a coke plant that was later indicted, convicted, and ordered to pay fines and conduct community impact studies.[201] Other examples of citizen involvement in promoting enforcement include air quality sampling by groups such as the Global Community Monitor and the Louisiana Bucket Brigade.[202]

If agencies are to take advantage of these new streams of data, they will need to develop protocols for collecting, storing, processing, and using the information. EPA has already begun doing so. As of mid-2016, the agency had charged its National Advisory Council for Environmental Policy and Technology with the preparation of a report detailing how EPA can best take advantage of citizen science, including ways to ensure data quality and security.[203] An official in EPA's Office of Air and Radiation described the agency's long-term goal as "harmonization, a synthesis of the gold standard monitoring network [run by government] with the evolving sensor technology" being used by individuals and community groups.[204] The agency has created an "Air Sensor Toolbox for Citizen Scientists" to provide guidance on sampling methodologies, calibration and validation approaches, measurement methods options, data interpretation, and low-cost sensor performance.[205] EPA has hosted training workshops,[206] described its research on air sensor monitoring and analysis technologies,[207] provided information about a project designed to demonstrate the capabilities of new real-time monitoring technology (called the Village Green Project),[208] and created a web-based tool (called Real-Time Geospatial Data Viewer) that can show air quality data collected by individuals while walking, biking, or driving.[209]

State environmental officials have undertaken similar initiatives. As indicated above,[210] Virginia environmental officials developed criteria for the

---

201. *See* Rejeski & McElfish, *supra* note 67, at 63.

202. *See* ELI, CLEARING THE PATH, *supra* note 76, at 25.

203. NATIONAL ADVISORY COUNCIL FOR ENVIRONMENTAL POLICY AND TECHNOLOGY, EPA STRATEGIC DIRECTIONS ON USING CITIZEN SCIENCE FOR ENVIRONMENTAL PROTECTION 2 (2015), https://www.epa.gov/sites/production/files/2015-09/documents/nacept_charge_on_citizen_science_final.pdf.

204. Yardley, *supra* note 192.

205. Ron Williams et al., *EPA's Air Sensor Toolbox for Citizen Scientists*, EPA, https://www.epa.gov/sites/production/files/2015-12/documents/r7tools_sensors_citizenscience_poster_508.pdf (last visited Jan. 25, 2017).

206. *See Community Air Monitoring Training: A Glimpse into EPA's Air Sensor Toolbox*, HEALTH & ENVTL. FUNDERS NETWORK, http://www.hefn.org/connect/event/community_air_monitoring_training_a_glimpse_into_epas_air_sensor_toolbox (last visited Jan. 8, 2017).

207. *See Air Monitoring, Measuring, and Emissions Research*, EPA, https://www.epa.gov/air-research/air-monitoring-measuring-and-emissions-research#ngamt (last updated Dec. 20, 2016).

208. *Village Green Project*, EPA, https://www.epa.gov/air-research/village-green-project (last updated Jan. 5, 2017).

209. *Real Time Geospatial Data Viewer (RETIGO)*, EPA, https://www.epa.gov/hesc/real-time-geospatial-data-viewer-retigo (last updated Sep. 8, 2016).

210. *See supra* note 76 and accompanying text.

appropriate use of data provided by nongovernmental monitors, including the identification of waters for follow-up monitoring by the state agency.[211] A partnership between the Wisconsin Department of Natural Resources and the University of Wisconsin–Cooperative Extension has created a program to train those who want to volunteer to participate in stream water quality monitoring efforts.[212] The initiatives described in this subpart merely scratch the surface of the potential to use data generated by community groups and individuals with access to new and cheaper information technologies to bolster compliance and enforcement.

D. *Information Technology Advances and Enhanced Transparency*

Transparency is a critical value of democratic government; one scholar identified it as "clearly among the pantheon of great political virtues."[213] Among other things, transparency informs citizens of what their government is doing, thereby providing them with a base of information they can use to better participate in governance efforts.[214] Increased transparency has long been a goal of government officials, though progress in achieving it has been uneven.[215] Indeed, it is one of the five key elements of Next Gen.[216]

The new information technology that is available to agencies and nongovernmental entities has the potential to promote that goal.[217] Interested

---

211. *Citizen Water Quality Data*, *supra* note 67; *Citizen Monitoring Guidance*, VA. DEP'T OF ENVTL. QUALITY, http://www.deq.virginia.gov/Programs/Water/WaterQualityInformationTMDLs/WaterQualityMonitoring/CitizenMonitoring/Guidance.aspx (last visited Jan. 8, 2017); *see also* Hindin & Silberman, *supra* note 45, at 121 (describing Virginia's citizen water quality monitoring program as an effective example of citizen participation in environmental monitoring).

212. *Water Action Volunteers Stream Monitoring Program*, WATER ACTION VOLUNTEERS, http://watermonitoring.uwex.edu/wav/ (last visited Jan. 8, 2017). The Riverkeeper network also engages volunteers to help monitor surface water quality. *See, e.g.*, *What Is a Riverkeeper?*, POTOMAC RIVERKEEPER NETWORK, http://www.potomacriverkeepernetwork.org/riverkeepers/ (last visited Jan. 8, 2017).

213. Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 888 (2006).

214. As one of his first acts as President, President Obama issued a memorandum stating that: "Transparency promotes accountability and provides information for citizens about what their Government is doing. . . . Government should be participatory. Public engagement enhances the Government's effectiveness and improves the quality of its decisions." Transparency and Open Government, 74 Fed. Reg. 4685, 4685 (Jan. 21, 2009). Some have identified limits the Obama Administration has placed on transparency and have criticized the Administration for its performance in this arena. *See, e.g.*, Cary Coglianese, *The Transparency President? The Obama Administration and Open Government*, 22 GOVERNANCE 529, 536–41 (2009); Jason Ross Arnold, *Has Obama Delivered the 'Most Transparent' Administration in History?*, WASH. POST (March 16, 2015), https://www.washingtonpost.com/news/monkey-cage/wp/2015/03/16/has-obama-delivered-the-most-transparent-administration-in-history/?utm_term=.a774219e562c.

215. *See, e.g.*, David Markell, *"Slack" in the Administrative State and its Implications for Governance: The Issue of Accountability*, 84 OR. L. REV. 1, 66 (2005) (discussing transparency as a long-standing goal of the federal government and of EPA and identifying limitations).

216. *See Next Generation Compliance*, EPA, http://www2.epa.gov/compliance/next-generation-compliance (last updated Dec. 23, 2016).

217. *See. e.g.*, Making Open and Machine Readable the New Default for Government Information, 78 Fed. Reg. 28,111 (May 14, 2013).

citizens increasingly will have the capacity to obtain data themselves because of the diminishing cost and greater availability of monitoring equipment. Regulated parties are also increasingly providing monitoring information directly to the public by posting it on websites and through other initiatives.[218] In addition, the emergence of technological innovations in the fields of monitoring and reporting has the potential to transform the agency's capacity to serve as a clearinghouse of information, one of EPA's long-standing functions.[219] ECHO, summarized above, is a prominent example of EPA's efforts to serve in this capacity in the compliance realm by providing an online platform for making compliance data easily accessible, including to regulated parties who will be able to assess relative performance, and communities who will be able to evaluate environmental compliance more effectively and pressure government and regulated entities alike to address problems.[220]

Newly available and more easily accessible information is likely to have a variety of effects on the roles of citizens, regulated parties, and government. According to EPA enforcement officials, transparent treatment of the new information stemming from advanced emissions monitoring and e-reporting can empower "communities and the marketplace to play a more active role in compliance oversight and improve the performance of both the government and regulated entities" by "provid[ing] more accurate, complete, and timely information on pollution sources, pollution, and compliance."[221] For example, additional regulated party-generated data, in tandem with citizen science, will yield improved insights about both absolute and relative performance, educating all interested stakeholders concerning the performance of different members of the regulated community.[222] That information is likely to make it easier for citizens to initiate enforcement actions and more likely they will do so. At the same time, the government will need to be mindful of limits on citizen capacity to produce reliable and salient data and to use it as EPA believes is appropriate.[223] Such a rise in citizen enforcement may increase the need for coordination of enforcement efforts.[224]

---

218. *See* Markell & Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming), *supra* note 29.

219. Hannah J. Wiseman, *Regulatory Islands*, 89 N.Y.U. L. REV. 1661, 1672 (2014) (discussing information deficits as a shortcoming in the use of states as laboratories of democracy and suggesting a stronger federal role as a clearinghouse); Markell, *supra* note 215, at 33–34; David L. Markell, *The Federal Superfund Program: Proposals for Strengthening the Federal/State Relationship*, 18 WM & MARY J. OF ENVTL. L. 1, 77 n.200 (1993) (discussing the federal government's potential to serve as a clearinghouse).

220. *See supra* notes 125–133 and accompanying text (discussing EPA's identification of better transparency as a key aspect of ECHO). We do not mean to understate the extent of the challenge in obtaining and managing data. *See supra* note 219.

221. Hindin & Silberman, *supra* note 45, at 106.

222. *See* Giles, *supra* note 19, at 26.

223. *See* Thalia González & Giovanni Saarman, *Regulating Pollutants, Negative Externalities, and Good Neighbor Agreements: Who Bears the Burden of Protecting Communities?*, 41 ECOLOGY L.Q. 37,

There is also considerable potential for regulated parties to improve compliance performance because of their access to additional, and more timely information. Additional information about compliance concerns, or circumstances that may trigger such concerns, has the potential to empower regulated parties, in other words, to use that information to improve performance.[225] Information about relative performance also has the potential to put pressure on lower-performing companies to improve their performance.[226]

Another benefit of transparency involves its potential to reshape relationships between regulated parties and nearby communities in a cooperative way. As the Assistant Administrator for EPA's Office of Enforcement and Compliance Assurance has recognized, the emergence of these innovations may enable interactions that do not depend on EPA but instead may occur organically within civil society (for example, between regulated parties and community groups, among regulated parties, or among community groups).[227] Uninformed communities and citizen groups are likely to lack both the incentive and the capacity to interact constructively with regulated entities. Those well equipped with information made available on governmental platforms will be in much better positions to engage regulated entities.[228]

Beyond its impact on the roles citizens and regulated parties may play, the increased information about noncompliance may assist EPA in making decisions about how to allocate its enforcement resources. Resource constraints have made it difficult to sustain the kind of enforcement presence the agency has traditionally had.[229] EPA's need to hire analytics experts and invest in new technological capacity may exacerbate these resource constraints.[230] EPA is

40–41 (2014) (discussing a case study in which even citizens in an affluent community were overwhelmed by governance duties).

224. *See* David Freeman Engstrom, *Agencies as Litigation Gatekeepers*, 123 YALE L.J. 616, 621 (2013) (discussing the role of agencies as "gatekeepers" of private enforcement activity). Citizen-generated data, in particular, at least in the view of some commentators, has the potential to "empower the public by giving them ownership over the data collection process, and teach data literacy through the act of collection." Williams, *supra* note 109, at 8.

225. *See* Sarah E. Light & Michael P. Vandenbergh, *Private Environmental Governance*, *in* ELGAR ENCYCLOPEDIA OF ENVIRONMENTAL LAW: DECISION MAKING IN ENVIRONMENTAL LAW 253 (LeRoy C. Paddock, Robert L. Glicksman & Nicholas S. Bryner eds., 2016). While new information has the potential to help regulated parties improve compliance, access to such information is a necessary, but not necessarily sufficient innovation to facilitate improved compliance.

226. Giles, *supra* note 19, at 25–26.

227. *Id.* at 24 (noting that new data may facilitate interactions between regulated parties and nearby communities).

228. *See Dynamic Governance, Part I*, *supra* note 29, at 623–25.

229. *See, e.g.*, Joel A. Mintz, *"Running on Fumes": The Development of New EPA Regulations in an Era of Scarcity*, 46 ENVTL. L. REP. 10,510, 10,511 (2016).

230. For a discussion of the increased costs that reliance on machine learning may generate, see *supra* note 101 and accompanying text.

likely to view citizen science as a means of easing resource concerns.[231] EPA will need to recognize the limits on citizen capacity to both produce reliable data and use it to pursue citizen enforcement actions.[232] Even if EPA determines that the information about noncompliance made possible by the introduction of e-reporting and other new technologies is accurate, it is likely to have to prioritize the problems it chooses to tackle and maintain appropriate levels of enforcement capacity internally.[233]

A final possible impact of increased and more easily accessible data that we briefly highlight is its potential to spur follow-up efforts by the government and others to learn more about risks and possible sources of pollution.[234] The discovery that ambient conditions in a particular location are unhealthy may spur those who live, work, or recreate there to monitor pollution emitted by potentially responsible sources. EPA may also use that information to engage in its own information-gathering activity to verify noncompliance and compile evidence for use in an enforcement action.[235] In some cases, new data or newly organized data may reveal regulatory gaps that reflect inadequate regulation of upwind or upstream sources and that need to be plugged. EPA has used data in the Toxic Release Inventory compiled under the Emergency Planning and Community Right-to-Know Act,[236] for example, to not only monitor compliance with existing regulatory standards and identify enforcement

231. Some suspect that "[o]ne of the primary motivations for the EPA to involve private parties in environmental enforcement has been a steadily declining level of enforcement resources." Sarah L. Stafford, *Private Policing of Environmental Performance: Does It Further Public Goals?*, 39 B.C. ENVTL. AFF. L. REV. 73, 74 (2012). At least in theory, advances in information technology will enable governments to "better . . . target limited compliance and enforcement resources on remaining pollution and noncompliance problems." Hindin & Silberman, *supra* note 45, at 106.

232. *See* Thalia González & Giovanni Saarman, *Regulating Pollutants, Negative Externalities, and Good Neighbor Agreements: Who Bears the Burden of Protecting Communities?*, 41 ECOLOGY L.Q. 37, 40–41 (2014) (discussing a case study in which even citizens in an affluent community were overwhelmed by governance duties).

233. *See* Markell & Glicksman, *Unraveling the Administrative State: Mechanism Choice, Key Actors, and Policy Implementation Tools* (forthcoming), *supra* note 29 (noting that EPA's awareness of noncompliance of nonmajor NPDES permittees is likely to increase as e-reporting begins under the 2015 e-reporting rule).

234. Our list in the text of possible impacts is intended to be illustrative, not comprehensive.

235. *See* Bass, *supra* note 77, at 19 ("With today's analytical tools, the overwhelming amount of data real time monitoring would create is now manageable, and new dissemination tools would make it possible to share such data publicly."); *see also supra* notes 199-200 and accompanying text (discussing EPA's follow-up investigations concerning Volkswagen's defeat devices); Macey, *supra* note 63, at 1630 ("The public has unprecedented means of generating data, aided by wireless sensor networks, personal exposure assessments that peer inside unregulated spaces such as the home and human body, and peer-to-peer data sharing."); *cf.* Scott Burris, *Public Health Law Monitoring and Evaluation in a Big Data Future*, 11 I/S: J.L. & POL'Y FOR INFO. SOC'Y 115, 118 (2015) (arguing that "new methods of analysis suited to big data may/should allow us in time to deal better with common variation in enforcement").

236. *See* 42 U.S.C. § 11023 (2012); *Toxic Release Inventory (TRI) Program*, EPA, https://www.epa.gov/toxics-release-inventory-tri-program (last updated Jan. 6, 2017).

priorities, but also "to help assess whether new regulations are needed to address environmental problems."[237]

Critically, as EPA has acknowledged, information must be correct for transparency to work effectively in these ways.[238] EPA's aspiration is that the monitoring, reporting, and transparency components of Next Gen will ultimately work synergistically to increase the chances that errors are not made.[239] As we have suggested above,[240] work remains to be done in this arena, but progress is being made.

CONCLUSION

Some have argued that the construction of the federal environmental statutes and the realities of modern technology are fundamentally mismatched.[241] Professor Macey, for example, contends that "[a]s environmental law evolves from a data-poor to data-intensive enterprise, the study of pollution control and ecosystem management will have to respond."[242] Further, he adds that "the conversion of data into useful, policy-relevant knowledge will change dramatically," replacing "the 'architecture of ignorance' that is currently in place."[243]

This Article considers how technological advances that promise to yield expanded volumes of data and enhanced capacity to mine it have the potential to shape governance efforts, with a special focus on the compliance realm. The Article demonstrates that such technological advances—especially new and

237. John D. Echeverria & Julie B. Kaplan, *Poisonous Procedural "Reform": In Defense of Environmental Right-to-Know*, 12 KAN. J.L. & PUB. POL'Y 579, 583 (2003).

238. Giles, *supra* note 19, at 26; Katrina Fischer Kuh & David L. Markell, *Informational Regulation, the Environment, and the Public* (forthcoming 2017) (reviewing several informational regulation strategies and identifying accuracy of information as an important element); *see also supra* Part I.B.1 (discussing data quality challenges).

239. Giles, *supra* note 19, at 26 ("Next Gen principles for advanced monitoring and electronic reporting go hand in hand with transparency: providing accurate information on real pollution issues.").

240. *See supra* Part I.B.1-2.

241. *See, e.g.*, Macey, *supra* note 63, at 1630 ("[E]nvironmental law is surrounded by an architecture that only makes sense in a world where data are scarce. It is constructed as statutes are stretched to accommodate spatial and temporal gaps in understanding. We gather data at broad spatial scales rather than along streetscapes, within neighborhoods, or in other realms of individual experience."). New information technologies have the capacity to address such mismatches. *See* Kennedy, *supra* note 32, at 126 ("For regulators, precise information can help create more specialised, decentralised, and sophisticated organizations. Quantification and visualization can better communicate environmental problems. Closer identification of problems allows policymakers to match the scale of the problem with the appropriate scale of response.").

242. Macey, *supra* note 63, at 1631.

243. *Id.*; *see also id.* at 1641 ("The laws are designed to make decisions in a data-poor context: based on data that agencies do not have (and firms might be in a better position to provide), with regulatory responses that occur despite what agencies do not know.").

improved monitoring capacity, advances in information dissemination, and improved data analysis—have significant potential to transform governance efforts to promote compliance. Such transformation is likely to affect not only the "how" of compliance promotion (the legal mechanisms and tools used to promote compliance), but also the "who" (the identity of the actors engaged in different ways in promoting compliance). The Article identifies some of the potential benefits of these transformative developments, as well as some of the challenges, and grounds the assessment by considering these issues in the context of EPA's ongoing efforts.

To close on a positive note, if EPA thoughtfully tackles the challenges that reliance on new data streams poses, the prospects for success of its effort to transform its enforcement and compliance programs should improve. Without discounting the challenges, including those we have reviewed above, some of the signs thus far are promising. For example, the use of advanced technology such as electronic reporting has already resulted in more accurate reporting on compliance status in states like Ohio. EPA has made electronic reporting its default position and has required it for water pollution and hazardous waste regulatory programs. The improved accuracy of compliance-related data and its improved availability for data mining and other forms of sophisticated analysis should enable EPA to work more effectively with its state partners to identify and target noncompliance problems and thereby foster higher compliance rates. The potential benefits of better integration of civil society into agency enforcement activities are reflected in a myriad of examples. These include instances in which community groups and environmental activists generate information for EPA. They also include contexts in which information is made available by the agency through public platforms such as websites to publicize poor performance by regulated entities and resulting environmental threats. As experience with the new technologies grows, EPA's continued investment of the necessary resources, including types of staff expertise, will enable the agency to continue to make progress in tackling challenges relating to data collection and analysis and thus improve regulatory compliance.