2016

Technological Innovation, Data Analytics, and Environmental Enforcement

Robert L. Glicksman
George Washington University Law School, rglicksman@law.gwu.edu

Recommended Citation

Glicksman, Robert L. and Markell, David L. and Monteleoni, Claire, Technological Innovation, Data Analytics, and Environmental Enforcement (September 19, 2016). Ecology Law Quarterly, Vol. 44, No. 1; FSU College of Law, Public Law Research Paper No. 815; GWU Law School Public Law Research Paper No. 2016-48; GWU Legal Studies Research Paper No. 2016-48. Available at SSRN: http://ssrn.com/abstract=2840944

DRAFT—PLEASE DO NOT CITE WITHOUT AUTHOR PERMISSION
TO BE PUBLISHED IN VOLUME 44, ISSUE # 1 OF THE ECOLOGY LAW QUARTERLY

TECHNOLOGICAL INNOVATION, DATA ANALYTICS, AND ENVIRONMENTAL ENFORCEMENT

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

Abstract

Technical innovation is ubiquitous in contemporary society and contributes to its extraordinarily dynamic character. Sometimes these innovations have significant effects on the state of the environment or on human health and they have stimulated 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." 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 (EPA)'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 EPA need to be aware of in opening this emerging bundle of new tools and making use of them to address real-world environmental needs.

Table of Contents

I. INTRODUCTION

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. The automobile, which transformed society economically and socially, was made

---

* 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.
** 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.

possible by the development of the internal combustion engine.[1] A side effect of this transformative technological innovation is its generation of air pollution, which society determined required regulatory attention under the 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, it assumed that the automobile industry would use a second category of technologies, catalytic converters, to limit the harmful effects of the initial technological innovation (the internal combustion engine).[6] Regulation has evolved over time as control technologies have improved,[7] efforts to control one pollutant affected the feasibility of controlling others, increases in fuel efficiency affected emission control performance, and persistent air pollution problems, especially in urban areas, triggered the need for more effective pollution controls.[8]

---

[1] "Perhaps no invention affected American everyday life in the 20th century more than the automobile." *U.S. History, Pre-Columbian to the New Millennium: The Age of the Automobile*, http://www.ushistory.org/us/46a.asp. 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, Natalie Messer Betts & John D. Graham, *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 [Environmental Protection Agency] to order refiners to alter gasoline formulations to protect the catalytic converters").

[7] 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).

[8] *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

As the development of motor vehicle emissions standards demonstrates, Congress, in selecting approaches from its regulatory tool box[9] has frequently required that regulated entities comply with standards that specify levels of performance that are achievable using the best available technology for the industry concerned.[10] Some environmental laws 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 have been successful, leading to the discovery and implementation of new technologies that limit or avoid pollution,[12] including the motor vehicle emission standards that helped prompt the development of the catalytic converter.[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]

---

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 (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, e.g.*, 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").

[9] *See generally* OFFICE OF TECHNOLOGY ASSESSMENT, ENVIRONMENTAL POLICY TOOLS: A USER'S GUIDE, OTA-ENV-634 (1995), https://www.princeton.edu/~ota/disk1/1995/9517/9517.pdf.

[10] *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) (CAA).

[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] *See* Thomas O. McGarity, *Radical Technology-Forcing in Environmental Regulation*, 27 LOY. L.A. L. REV. 943 (1994) (citing as examples of successful statutory technology-forcing ventures the phaseout of the pesticide Mirex and the phaseout 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 phaseout)

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

[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

For some processes, the search for technological fixes is longstanding, and 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 (GHGs) 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 its energy mix.[17]

In addition to technologies that create environmental concerns and others that can help to ameliorate their adverse effects, another category of technologies that are of foundational importance to environmental law are those that monitor environmental conditions. Some monitoring tools measure ambient environmental quality or otherwise have the potential to shed light on the state of the environment, while others track releases of pollution (or other indicia of risk associated with regulated party operations). Monitoring of both forms is indispensable to assuring (or at least promoting) compliance with environmental norms. Compliance is central to the success of environmental law; it is impossible to achieve a regulatory goal if the means adopted to promote it are ignored.[18] For the same reason, a vigorous enforcement 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

---

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 (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 "Federal 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, 115 (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 "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

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 those involved in, affected by, and responsible for regulating activities subject to environmental regulation to identify, measure, share, analyze, report on, and respond to the effects of those activities. 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] In the environmental compliance sphere, these technological advances have the potential to transform the capacities of the suite of actors involved, including government officials, regulated parties, and interested citizens.[27] They also have the potential to transform relationships between and among these actors, and the roles each performs.[28]

---

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 supra* § IIA.

[23] Abraham R. Wagner & Paul Finkelman, *Security, Privacy, and Technology Development: The Impact on National Securit*y, 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 (2014).

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

[26] *See, e.g.*, Joseph Jerome, *Big Data: Catalyst for A Privacy Conversation*, 48 IND. L. REV. 213 (2014). 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., Springer) (forthcoming), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2831040.

[27] 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. 10984 (2014).

[28] *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, *Dynamic Governance in Theory and Application, Part II* (forthcoming) [hereinafter *Dynamic Governance, Part II*]; David L. Markell & Robert L. Glicksman, *Dynamic Governance in Theory and Application, Part I*, 58 ARIZ. L. REV. 563 , 618-29 (forthcoming) [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*].

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 policymaking and implementation process and thereby play a significant role in transforming the governance landscape. Information technology can increase the efficiency and effectiveness of the U.S. Environmental Protection Agency (EPA)'s compliance and enforcement programs (e.g., by providing relevant information at less cost and thereby enabling agencies to reduce the levels of staffing needed to catalog and process the information and reallocate resources to other areas of need); it can upgrade EPA-state relations (e.g., by fostering information sharing and better coordinated enforcement activity between EPA and its state partners under the environmental cooperative federalism statutes); it can improve regulated parties' capacity (e.g., to identify, diagnose, and address compliance concerns); and it can enhance community groups' ability to participate in governance efforts (e.g., 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. . . . "[29]

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 considering and addressing these challenges by reviewing an ongoing initiative by EPA called Next Generation Compliance (or Next Gen), which aims to "transform" traditional environmental enforcement practices at the agency, largely but not entirely through greater reliance on advanced monitoring and reporting technologies.[30] The success or failure of that effort is likely to be shaped by the agency's ability to exploit emerging technologies and recognize and respond effectively to challenges in doing so.

Part II 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 II also identifies a series of significant challenges in using this information to

---

[29] Ronan 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). Kennedy contends that "no coherent perspectives, approaches, or frameworks have developed" on the relationships between information technology and environmental regulation. *Id*.

[30] 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.

promote improved environmental compliance and more effective enforcement. We categorize these challenges according to the activities that relate to the data – primarily, data collection and analysis.

In Part III, we analyze the ways in which the information technology revolution may influence environmental compliance and enforcement. Section A discusses ways in which EPA anticipates that the new data will contribute to improved compliance and enforcement, some of which EPA has already begun pursuing in connection with its Next Gen initiative. Section B covers the use of new information technologies by regulated entities, while section 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, as well as 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.

II. 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.[31] Computers and devices linked to them, such as sensors of various types, are capable of producing data in settings of previous data scarcity[32] and in volumes that dwarf previously available information. These same devices, or others to which they are connected, are capable of analyzing the data that is captured more quickly and thoroughly than ever before.

A. A Threshold Challenge: Defining the Key Terms

[31] See Liane Colonna, *A Taxonomy and Classification of Data Mining*, 16 SMU SCI. & TECH. L. REV. 309, 369 (2013) (citing The Technology Review Ten, *MIT Tech. Rev.*, Jan.-Feb. 2001, at 97) (listing "data mining" as "one of the ten emerging technologies that would change the world"). 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.

[32] *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), http:// www.foreignaffairs.com/articles/139104/kenneth-neil-cukier-and-viktormayerschoenberger/the-rise-of-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.'").

Advances in information technology have prompted a new vocabulary that includes terms such as "data mining" and "big data."[33] The use of terms such as "big data," a term coined in 1997,[34] tends to be context-specific. As a report to the President noted in 2014:

> 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.[35]

A widely used definition centers on the three "Vs" – "high-volume, -velocity and -variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision making."[36] Big data is thus a term that has been used to refer to large volumes of information,[37] the techniques used to generate and disseminate it, and the

---

[33] 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, Issue 3 (2015); *Symposium: Big Data Future Part Two*, 11 I/S: J. L. & SOC'Y FOR INFO. SOC'Y, Issue 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).

[34] Breggin & Amsalem, *supra* note 27, at 10984 ("The term 'big data' was first used in 1997, marking the increasing trend of rapid data growth.").

[35] EXECUTIVE 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 & Dianne Callan, *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").

[36] 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, 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* not3 27, at 10984 (citing Gartner, Inc., *Big Data*, IT Glossary, http://www.gartner.com/it-glossary/big-data/); SEIZING OPPORTUNITIES, *supra* note 35, 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. e.g.*, 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, 14 (2015) (quoting ORG. FOR ECON. CO-OPERATION & DEV., SUPPORTING INVESTMENT IN KNOWLEDGE CAPITAL, GROWTH AND INNOVATION 12 (2013)) ("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.").

[37] *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

methods used to analyze it.[38] Groups like the U.S. National Institute of Standards and Technology, the International Standards Organization, and the W3 have made efforts to develop a common set of big data definitions, taxonomies, formats, and reference architectures.[39]

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,[40] and fact-based management" are extensively used "to drive decisions and add value."[41] Our purpose is to highlight the promise provided by the emergence of new forms of data generation and analytical capacity to agency functions such as regulatory enforcement and 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 III explains, EPA has or soon will have vast troves of new data at its disposal. The data are being generated by the agency itself, by regulated entities (often under EPA mandate or

---

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 [hereinafter ELI, BIG DATA] ("'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.").

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 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., June 2011)).

[38] *See* Terry, *supra* note 37, 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* Breggin & Amsalem, *supra* note 27, at 10985 ("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").

[39] Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 545-46 (2014).

[40] "New machine learning techniques outperform traditional regression approaches to prediction." Anthony J. Casey & Anthony Niblett, *Self-Driving Laws* at 4, http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2804674. [hereinafter Casey & Niblett, *Self-Driving Laws*]; *see also* Anthony J. Casey & Anthony Niblett, *The Death of Rules and Standards* 2-3, http://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*].

[41] Borden & Baron, *supra* note 32, at 23 (quoting THOMAS H. DAVENPORT & JINHO KIM, KEEPING UP WITH THE QUANTS: YOUR GUIDE TO UNDERSTANDING AND USING ANALYTICS 3 (2013)). *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'").

encouragement to do so), by third-party auditors,[42] and by the broader civil society, including environmental non-governmental organizations and community groups. A critical question, at least for our purposes, is what can and will EPA do with the 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 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'"[43] both in individual criminal[44] and business regulatory contexts.[45] The questions addressed in the next section involve what challenges EPA, or any organization seeking to rely on new or enhanced data streams, and improved capacity to mine that data, to bolster compliance and enforcement, will likely face in doing so, and how those challenges might be met in ways that maximize the value of new information technologies.

B. Systemic Challenges in the Use of Data Analytics

The revolution in information technology has the potential to improve understanding of the state of environmental compliance and thereby to improve compliance through the combined efforts of government, regulated entities, and civil society. It has already begun and will continue to generate enormous quantities of new information about environmental conditions and

---

[42] "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." David A. Hindin & Jon D. Silberman, *Designing More Effective Rules and Permits*, 7 GEO. WASH. J. OF ENERGY & ENVTL. L. 103, 113 (2016).
[43] Ismail Cem Kuru, *Your Hard Drive Is Almost Full: How Much Data Can the Fourth Amendment Hold*, U. ILL. J.L. TECH. & POL'Y, Spring 2016, at 89, 92.
[44] *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) (noting that "law 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) (describing how the collection of " massive amounts of surveillance camera data and mining it for law enforcement purposes . . . promises to reduce crime and increase personal safety").
[45] Joy Heath, *Government Highlights New Focus on Physician Fraud*, HEALTH LAW., Aug. 2015, at 36, 38 n.1 (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"); David Gray et. al., *Fighting Cybercrime After United States v. Jones*, 103 J. CRIM. L. & CRIMINOLOGY 745, 798 (2013) (data analytics and cybercrime). The Securities and Exchange Commission (SEC) has created an entire division, the Division of Economic and Risk Analysis (DERA), "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." Jena Martin, *Changing the Rules of the Game: Beyond A Disclosure Framework for Securities Regulation*, 118 W. VA. L. REV. 59, 93 (2015); *cf*. Rachel E. Barkow, *The New Policing of Business Crime*, 37 SEATTLE U. L. REV. 435, 450 (2014) (referring to the SEC's creation of DERA in 2009 as a "think tank" to "'integrate financial economics and rigorous data analytics into the core mission of the SEC' and across all of the SEC's activities, from rulemaking to enforcement and examination"); Henry T. C. Hu, Keynote Address: *The SEC, Dodd-Frank, and Modern Capital Markets*, 7 N.Y.U. J.L. & BUS. 427, 435-36 (2011) (describing the use of data analytics in securities regulation enforcement). The SEC's website describes DERA's functions. Securities and Exchange Comm'n, *Economic and Risk Analysis*, https://www.sec.gov/dera.

performance, facilitate the distribution of that information, and provide new tools for analyzing the data, and adjusting public and private decisions based on the results. But revolutions do not always occur seamlessly. Instead, new opportunities are not always seen or seized, wrong turns are taken, and unintended consequences occur. Efforts to incorporate data newly available because of technological advances and new analytical techniques, including but not limited to predictive algorithms, into environmental governance are likely to produce all three. The transformation of EPA's compliance and enforcement programs through increased reliance on emerging improvements in capacity to collect and analyze data may not be smooth. To ease the transition, EPA should take steps to prepare for, and respond to a series of significant challenges presented by, the use of new data sources as a tool in enforcement-related decisionmaking. In this part, we identify several such challenges that relate to data collection, dissemination, and analysis.[46]

1. *Data Collection, Storage, and Transport*

The data collection challenges facing policymakers 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; weeding out poor quality or unreliable data resulting from poor equipment or lack of training in the use of equipment; aggregation of information of differing levels of quality from sources of differing levels of reliability as a result of variations in the quality of the equipment used and in the levels of training and experience in those operating it; adjusting verification and quality control requirements to different intended uses of the data; and protecting against hacking of computer systems that store the data resulting in data corruption.[47] Informed policy-based decisions based on any kind of information depend on the quality of the data.[48] 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-

---

[46] 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 29, 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").

[47] *See, e.g.*, 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, 26 and accompanying text. For discussion of some of the legal and policy challenges, see, e.g., Mattioli, *supra* note 39, at 536 (discussing how 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 (security and privacy challenges); Margaret Hu, *Big Data Blacklisting*, 67 FLA. L. REV. 1735 (2015) (constitutional challenges); Hu, *supra* note 36, at 785 (discussing "challenges of big data-driven national security policymaking and the role of big data cybersurveillance in national security law").

[48] "[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 27, at 10992.

governed quality data."[49] Poor data quality can adversely affect decisions by individuals, firms, and governments.[50] There is a long history of data accuracy problems in EPA compliance-related data.[51] Data that continue to be afflicted by errors or inaccuracies 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) or the best solutions to address those problems.[52] 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.[53]

Data quality problems are likely to include both incomplete and inaccurate data.[54] 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 transformative 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,"[55] although some community groups have sought to monitor in areas traditionally neglected by

---

[49] 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). 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.").

[50] 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 29, at 136.

[51] *See, e.g.*, *Dynamic Governance, Part I*, *supra* note 28 (forthcoming) (Part IIA.1); Markell & Glicksman, *A Holistic Look*, *supra* note 28, at 47-48 (describing inaccuracy and incompleteness of data on compliance and enforcement).

[52] Price, *supra* note 47, at 1414.

[53] *See infra* § IIIB.2.b.

[54] Graves et. al., *supra* note 50, at 121-31 (reviewing data quality problems in developing national security policies).

[55] 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).

government monitors.[56] 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.

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 whether the devices used to generate it are of high quality. 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."[57] 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.[58] EPA has described calibration and its importance to environmental monitoring:

> 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.[59]

The challenge of ensuring proper calibration is obviously going to be much greater if the devices generating enforcement-related data are being operated by myriad non-governmental sources.

---

[56] 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.").

[57] *See* Douglas Main, *Your Office Air Is Killing You*, NEWSWEEK, June 2, 2016, http://www.newsweek.com/2016/06/10/indoor-air-pollution-revolution-465531.html; *see also* Patrick Ambrosio, *Low-Cost 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 IIIC.1.

[58] *See, e.g.*, Hindin & Silberman, *supra* note 42, at 111 (explaining that immediate feedback technology, which supply 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").

[59] U.S. EPA, AIR SENSOR GUIDEBOOK, EPA 600R-14-159 (June 2014), at 21, 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. U.S. EPA, NEXT GENERATION AIR MONITORING (NGAM) VOC SENSOR EVALUATION REPORT, EPA/600/R-15/122 (May 2015), https://cfpub.epa.gov/si/si_public_record_report.cfm?dirEntryId=308114&simpleSearch=1&searchAll=Next+generation+air+monitoring.

Lack of sophistication, experience and training in using the equipment used collect data also are likely to produce transmission errors.

Data such as environmental monitoring information may be used for different purposes, such as ascertaining the need for more stringent emission control standards, devising strategies to improve ambient conditions, or identifying noncompliance and collecting evidence to support enforcement action. As a result, regulators are likely to 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."[60]

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 correct variables – those that are being monitored for compliance and from which one can infer information about those variables. Further, the data must be collected at appropriate locations. For example, as the discussion in Part III 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, for a variety of reasons.[61] If a purpose of such monitoring is to bolster the capacity to identify noncompliance, data gathering efforts must be designed with that objective in mind.[62]

Data collection therefore requires coordination among multiple data sources,[63] checks on quality, and compliance with legal and regulatory requirements.[64] Some see these data collection

---

[60] Daniel Rejeski & James McElfish, *Citizen Science*, 33 ENVTL. F. 62, 63 (July/Aug. 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." Va. Dep't of Envtl. Qual., *Levels of Citizen Water Quality Data in Virginia* 1, http://www.deq.virginia.gov/Portals/0/DEQ/Water/WaterQualityMonitoring/CitizenMonitoring/Data_use_form.pdf [hereinafter *Citizen Water Quality Data*].

[61] 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) [hereinafter McGarity, *Hot Spots*].

[62] Gregg P. Macey, *The Architecture of Ignorance*, 2013 UTAH L. REV. 1627, 1659 (2013).

[63] 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." Matthew Gordon, *Big Data: It's Not the Size That Matters*, 7 J. NAT'L SECURITY L. & POL'Y 311, 313 (2014); *see also* McGarity, *Hot Spots*, *supra* note 61, at 1481-83 (discussing limitations on the accuracy of modern, mobile monitoring technologies).

[64] W. Nicholson Price II, *Big Data, Patents, and the Future of Medicine*, 37 CARDOZO L. REV. 1401, 1415 (2016). 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

and transmission-related challenges as a relatively intractable problem, contending, for example, that relatively few analytics solutions "work robustly with multimodal and heterogeneous data types."[65] But even if the data collected and transmitted to EPA are initially of high quality, government databases may be susceptible to hacking by outsiders, which could result in disclosure of information the government regards as confidential[66] or corruption of stored data that impairs its utility in supporting enforcement action.[67] This problem is obviously not confined, however, to databases comprised of information supplied by those outside the government.

It will be important for EPA to develop protocols for the generation and collection of data it perceives as most useful to advancing its Next Gen goals. Both researchers and policymakers have grappled with these technical challenges. As noted above, several groups have undertaken efforts to develop common definitions and other key elements of protocols.[68] Environmental agencies, too, have 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 waters for future agency monitoring.[69] As one observer has noted, "providing data standards . . . is essential for meaningful data exchanges, which is a critical part of transparency and accountability."[70]

> 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.

Matthew Gordon, *Big Data: It's Not the Size That Matters*, 7 J. NAT'L SECURITY L. & POL'Y 311, 314-15 (2014). 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).

[65] 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).

[66] *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").

[67] *See, e.g.*, Chad Squitieri, Note, *Confronting Big Data: Applying the Confrontation Clause to Government Data Collection*, 101 VA. L. REV. 2011, 2027-28(2015) (discussing "the unique difficulties big data faces regarding storage. Data, including data stored in the 'cloud,' is susceptible to corruption while in storage. . . . Stored data is also susceptible to destruction."); *The Evolving Role of the Corporate Counsel: How Information Technology Is Reinventing Legal Practice*, 36 CAMPBELL L. REV. 383, 398 (2013) (remarks of Michael Rappa) ("We also have to worry about data corruption. . . . There are people with mal intent who are very interested in corrupting data . . . .").

[68] *See supra* note 39 and accompanying text.

[69] *Citizen Water Quality Data*, *supra* note 70. *See also* ENVIRONMENTAL 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 [Hereinafter ELI, CLEARING THE PATH](suggesting appropriate design considerations for projects to clear the path toward greater governmental access to, and reliance on, citizen science").

[70] 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).

2. *Analysis*

Data collection is merely the first step in the process of using new (and existing) information sources to improve environmental policy actions. EPA has long faced challenges in using its data efficiently and effectively.[71] Increasing volumes of data, from an increasing variety of sources, collected in an increasing variety of ways, will inevitably create new challenges, as well as new opportunities. Some large data sets present significant interpretive challenges.[72] A significant challenge for users of big data is designing methods for analyzing the data collected – "without the analytic ability to unlock key information and patterns, big data sets are of limited use."[73]

The three "Vs" that often characterize big data all present potential interpretive problems.[74] The data streams produced by new information technology can arrive quickly, in complex formats, and from a variety of sources, significantly contributing to "the challenge of finding signals in the noise"[75] and therefore of detecting useful patterns or relationships that are useful in making the decisions the data whose collection 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 . . .)."[76] Another analogized the analysis of big data to finding a needle in a haystack.[77]

---

[71] *See, e.g.*, Joanna Lau, *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").

[72] *See* FRANK PASQUALE, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION (2015); *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.").

[73] Breggin & Amsalem, *supra* note 27, at 10986; *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.).

[74] For discussion of the three Vs, see *supra* note 36 and accompanying text.

[75] Borden & Baron, *supra* note 32, at 21 (quoting BILL FRANKS, TAMING THE BIG DATA TIDAL WAVE: FINDING OPPORTUNITIES IN HUGE DATA STREAMS WITH ADVANCED ANALYTICS 5 (John Wiley & Sons, Inc. ed., 2012)).

[76] Fahey, *supra* note 37, at 325.

[77] Mattioli, *supra* note 39, at 557-58. Despite the challenges, techniques to find such needles exist and have been used with considerable success. *See, e.g.*, Yann LeCun, Yoshua Bengio & Geoffrey Hinton, *Deep Learning*, 521 NATURE 436 (May 28, 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).

The Program Director for Data Science at the National Institute of Standards and Technology 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 lack of understanding of what works in big data analytics algorithms;[78] lack of understanding of the foundational gaps in big data science; issues concerning evaluation methods, tools, and reference data; lack of understanding about the usability of big data systems and solutions; limited understanding of how the quality and context of input data affect derived conclusions; lack of multidimensional benchmarks for application to analytics tools and processes; and the need to determine how to evaluate which components are best suited to specific families of tasks.[79] 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.[80] For example, agencies are using tools such as a custom database created by IBM and sophisticated analytics to map interrelated criminal activity to draw connections between apparently unrelated cases involving illegal trafficking in hazardous substances or endangered species and share information with governments combatting eco-crimes,[81] and 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.[82] The challenges identified by the Program Director are 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 are to be realized.[83]

---

[78] According to one source, the "heart" of big data use is "the development of algorithms to find patterns in the data" so that relevant outcomes can be predicted and responses developed. Price, *supra* note 47, at 1415f. Among other things, pattern recognition processes "yield insights into individual behavior and characteristics." Sean Brian, Comment, *The Unexamined Life in the Era of Big Data: Toward A UDAAP for Data*, 40 U. DAYTON L. REV. 181, 183 (2015).

[79] Talukder, *supra* note 65, at 805-06.

[80] Breggin & Amsalem, *supra* note 27, at 10986; *see also* Ann R. Klee, *The Digital Transformation of Environment, Health, and Safety*, 33 ENVTL. F. 17, 17 (Mar.-Apr. 2016) (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 (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 serious problems, to identify criminal activity and to help us figure out where we should be focusing our time").

[81] Breggin & Amsalem, *supra* note 27, at 10987; David Maxwell Braun, *Big Data Analytics Helping to Protect Big Cats*, NAT'S GEOGRAPHIC, CAT WATCH, Feb. 12, 2013, http://voices.nationalgeographic.com/2013/02/12/big-data-and-analytics-helping-to-protect-big-cats/.

[82] Nat'l Oceanic and Atmospheric Adm'n, *Emergency Response Management Application (ERMA)*, http://response.restoration.noaa.gov/maps-and-spatial-data/environmental-response-management-application-erma.

[83] Statistical analysis of data uses samples to generalize about a larger population.

> More data lends greater predictive power to the generalizations that result from the study. . . . Increasing predictive power through increasing sample size is the driving force behind machine learning. 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.

One interpretive problem relevant to the environmental enforcement context may be connecting problematic environmental conditions to particular sources suspected of violating their individual emission limits. The question is whether it is possible to 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.[84] This is a problem of longstanding. Congress chose in 1972 to engraft a technology-based program for controlling point source discharges to surface water onto the pre-existing water quality-based regime in part because of the difficulty of identifying cause-and-effect relationships between discharges and ambient water quality problems. The difficulty of drawing that connection based on the scientific information then available hindered enforcement efforts by the states.[85]

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

Brian, *supra* note 78, at 183-84. "'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 following description:

> Artificially 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. In doing so, they amalgamate the wisdom of crowds.

Casey & Niblett, *Death of Rules*, *supra* note40, at 30.

But the amount of data available to decisionmakers 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 70, at 33 ("Sometimes the best way to hide key information is to bury it in massive datasets.").

[84] 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 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 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*, PROC. OF THE 2013 SIAM INT'L CONF. ON DATA MINING 467 (2013), http://www-bcf.usc.edu/~liu32/sdm_theory.pdf. *See also* McGarity, *Hot Spots*, *supra* note 61, 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.").

[85] Robert L. Glicksman & Mathew J. 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).

than before.[86] But new data streams do not inevitably facilitate making desired causal connections or otherwise producing the information needed to support policy decisions or provide the evidentiary foundations for enforcement actions. The data generated by newly available information technology may be comprised of a single aggregate signal, while the user may wish to separate the data into a mixture of unobserved component signals.[87] For example, smart meters can tell us the amount of energy used in a home at different parts of the day, but what may interest us is what individual appliances contributed to those patterns of energy usage. New algorithms for disaggregating such undifferentiated signals need to be and are being developed.[88] Similarly, the detection of chemical concentration limits above a regulatory threshold may be the trigger for investigation of potential emission violations by nearby sources. But if a chemical concentration exceeds a threshold at location A at time t, and there are multiple polluters in the relevant vicinity, 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. This causal attribution problem is not a new one; EPA and the states have long had difficulty tracing ambient water quality problems to responsible sources.[89] Although advanced information technologies may be able to alleviate the problem 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 "ensuri[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 bias."[90] A recent report to the President cautioned that

---

[86] *See* Kennedy, *supra* note 29, 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.").

[87] Matt Wytock & J. Zico Kolter, *Contextually Supervised Source Separation with Application to Energy Disaggregation* (Ass'n for the Advancement of Artificial Intelligence 2014), https://www.cs.cmu.edu/~mwytock/papers/context_full.pdf.

[88] *See id*. (suggesting contextual supervision as an alternative to the traditional approaches, supervised and unsupervised settings); *see also* J. Zico Kolter & Tommi Jaakola, *Approximate Inference in Additive Factorial HMMs with Application to Energy Disaggregation*, PROC. OF THE 15TH INT'L CONF. ON ARTIFICIAL INTELLIGENCE AND STATISTICS 1472 (2012), http://jmlr.csail.mit.edu/proceedings/papers/v22/zico12/zico12.pdf.

[89] Glicksman & Batzel, *supra* note 85, at 118-19.

[90] 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 27, at 10991 ("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 10994.

[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."[91] Similarly, a geographer 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."[92] 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."[93] The literature on machine learning, which includes a subfield on causality, is providing new techniques for addressing causation questions.[94] Use of those techniques has driven discoveries that promise increased understanding of problems such as climate change and more informed solutions.[95]

A perhaps apocryphal story about machine learning suggests the risks that may result from incomplete or flawed analysis of the information made available by new technologies. In an article entitled *Neural Network Follies*,[96] Neil Fraser describes a Pentagon effort to take advantage of computer technology to make armed forces tanks safer when in combat. The idea was to place a digital camera on each tank, connect each camera to a computer, and use the camera to constantly scan for outside threats such as an enemy tank hiding behind a tree. The computer would alert the tank crew to suspicious objects. To train the computer, the Pentagon used a "neural network." The research team took 100 photos of tanks behind trees and 100 photos of trees with no tanks behind them. Fraser describes the process used to "train" the network to identify suspicious activity as follows:

> The huge neural network was fed each photo one at a time and asked if there was a tank hiding behind the trees. Of course at the beginning its answers were completely random since the network didn't know what was going on or what it was supposed to do. But each

---

[91] SEIZING OPPORTUNITIES, *supra* note 35, at 7; *see also* Breggin & Amsalem, *supra* note 27, at 10944 (stating that "correlation does not necessarily indicate a cause-and-effect relationship").
[92] Miller, supra note 59, at 714.
[93] Breggin & Callan, *supra* note 35, at 121.
[94] *See, e.g.,* SUSAN ATHEY & GUIDO IMBENS, MACHINE LEARNING METHODS FOR CAUSAL EFFECTS, http://www.nasonline.org/programs/sackler-colloquia/documents/athey.pdf; Peter Spirtes, *Introduction to Causal Inference*, 11 J. MACHINE LEARNING RESEARCH 1643 (2010), http://www.jmlr.org/papers/volume11/spirtes10a/spirtes10a.pdf.
[95] *See, e.g.,* Yi Deng & Imme Ebert-Uphoff, *Weakening of Atmospheric Information Flow in a Warming Climate in the Community Climate System Model*, GEOPHYSICAL RESEARCH LTRS. 10.1002/2013GL058646 (2014), https://www.atmos.colostate.edu/~iebert/PAPERS/Deng_Ebert_Uphoff_InformationFlow_2014.pdf (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). 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, http://climatechange.cs.umn.edu/publications.php.
[96] Neil Fraser, *Neural Network Follies*, https://neil.fraser.name/writing/tank/ (Sept. 1998, updated February 2003).

> time it was fed a photo and it generated an answer, the scientists told it if it was right or wrong. . . . Over time it got better and better until eventually it was getting each photo correct. It could correctly determine if there was a tank hiding behind the trees in any one of the photos.[97]

But the Pentagon sought additional confirmation that the network could distinguish between pictures with tanks and those without. It "commissioned another set of photos (half with tanks and half without) and scanned them into the computer and through the neural network. The results were completely random. For a long time nobody could figure out why. After all nobody understood how the neural had trained itself."[98]

The explanation lay in a variable no one had considered, whether the sky was clear or cloudy when the picture was taken:

> Eventually someone noticed that in the original set of 200 photos, all the images with tanks had been taken on a cloudy day while all the images without tanks had been taken on a sunny day. The neural network had been asked to separate the two groups of photos and it had chosen the most obvious way to do it – not by looking for a camouflaged tank hiding behind a tree, but merely by looking at the colour of the sky. The military was now the proud owner of a multi-million dollar mainframe computer that could tell you if it was sunny or not.[99]

The take home message from this example is that data needs to be large and diverse enough to be representative of all salient values (e.g., here, the different weather conditions coupled with both the tank and the lack thereof) relating to the distribution that is the focus of the inquiry.

Further, databases often have systematic biases. An example from the environmental world is the higher ambient temperature that exists in urban areas and near buildings.[100] 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

---

[97] *Id*.
[98] *Id.*
[99] *Id*.
[100] *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.").

measurements background concentrations and amounts emitted by non-refinery sources or caused by fugitive emissions.[101] The challenge of ensuring that the results of monitoring activities by EPA, states, community groups, nongovernmental organizations, and individuals produce data whose analysis is capable of identifying facilities likely to be in noncompliance based on their proximity to problematic ambient conditions, which can then be further monitored to determine compliance status, is likely to be an ongoing one, however.[102]

III. 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 important innovations.[103] In the environmental enforcement context, these may include new and improved ways to identify noncompliance, prioritize the use of government investigation and enforcement resources (state

---

[101] Petroleum Refinery Sector Risk and Technology Review and New Source Performance Standards, 80 Fed. Reg. 75,177, 75,192 (Dec. 1, 2015).

[102] 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 it. "[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." Roslyn Fuller, *Structuring Big Data to Facilitate Democratic Participation in International Law*, 42 INT'L J. LEGAL INFO. 504, 505 (2014). Those who want to rely on big data as decisionmaking inputs need the physical infrastructure required to store the amount of data that needs to be collected, the resources (monetary and personnel) to collect, store, and analyze the data, and the expertise to perform these tasks.

In the private sector, firms often rely on technology startups who specialize in sifting through huge volumes of data. Mattioli, *supra* note 39, 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 47, 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.*, *See, e.g.*, Hoffman & Podgurski, *supra* note 90, 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 also* Cary Coglianese, *Optimizing Government for an Optimizing Economy*, *in* NEW ENTREPRENEURIAL GROWTH AGENDA __ (Dane Stangler ed., Ewing Marion Kauffman Found. 2016) (arguing that "the federal government's information technology infrastructure needs to rise to the task").

[103] Mattioli, *supra* note 39, at 543; *see also* SEIZING OPPORTUNITIES, *supra* note 35, at 61 ("Big data technologies are driving enormous innovation while raising novel privacy implications."); Byers, *supra* note 102, at 758 ("[B]ig data enables experimentation, often involving rigorous statistics analyses to identify what option is better.").

and federal) and enhance government capacity to pursue enforcement actions where appropriate, and engage both regulated entities and affected communities in seeking ways to foster better 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,[104] enforcement and compliance,[105] health and safety research,[106] interagency collaboration,[107] and public engagement."[108]

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.[109]

The potential value of technological innovations extends beyond informing government decisionmakers. For example, the information 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 information supplied by the new technologies can benefit "audiences inside and outside the policy arena."[110]

---

[104] *See, e.g.*, Breggin & Amsalem, *supra* note 27, at 10987 (referring to use of data supplied by WaterWatchers to prioritize improvements to city water infrastructure).
[105] *See, e.g.*, *id*. at 10990 (referring to use of meteorological and air quality data in real-time online supplied by the Village Green Project to advance air quality monitoring).
[106] *See, e.g.*, *id*. (referring to information supplied by the California Seafloor Mapping Program to improve maritime safety).
[107] *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 10987 (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).
[108] *Id*. at 10991; *id*. at 10990 (referring to interagency task force efforts to provide utility users access to their own energy data).
[109] Madison, *supra* note 73, 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*, *in* Jayasuriya & Ritscheske, *supra* note 35, at 3, 3 (quoting MANYIKA ET AL., *supra* note 37, 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.").
[110] 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).

Gregg Macey suggests that those who enacted our foundational environmental law infrastructure were "data-starved."[111] The hope is that "big data" and other technology-related innovations will increasingly lead to more informed evidence-based policymaking, and also to a more informed body politic.[112]

This Part explores some of the ways in which EPA is seeking to put technological innovations to work in its effort to improve its compliance and enforcement program. It focuses on EPA's Next Generation Compliance (Next Gen) initiative, which the agency initiated in 2013 and has touted as a transformative endeavor.[113] According to agency enforcement officials, "[t]oday's compliance challenges require a modern approach with new tools while continuing to employ vigorous enforcement as a backbone of environmental protection."[114] Next Gen is designed to provide those tools. As conceptualized by EPA,[115] Next Gen is comprised of five inter-related elements: regulation and permit design, advanced monitoring, electronic reporting, transparency, and innovative enforcement.[116] 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 agency's Next Gen 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. Section A, B, and C, respectively, highlight Next Gen's use of technological innovations by federal and state regulators, regulated entities, and civil society. Section D explores the potential of those innovations to increase the transparency of the activities of all stakeholders in the environmental regulatory process.

A. The Use of Information Technology by Regulatory Agencies

---

[111] Macey, *supra* note 57, at 1630.
[112] Carole Roan Gresenz, *Using Big Data to Assess Community Health & Inform Local Health Care Policymaking*, *in* Jayasuriya & Ritscheske, *supra* note 35 (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").
[113] EPA announced the new initiative in an article by EPA's Assistant Administrator for Enforcement, Cynthia Giles. Giles, *supra* note 19. For an early review of Next Gen, see David L. Markell & Robert L. Glicksman, *Next Generation Compliance*, 30 NATURAL RES. & ENV'T 22 (Winter 2016).
[114] Hindin & Silberman, *supra* note 42, at 103.
[115] For an alternative model for conceptualizing the design of a regulatory compliance and enforcement program, see *Dynamic Governance, Part I*, *supra* note 28.
[116] *See* U.S. EPA, *Next Generation Compliance*, http://www2.epa.gov/compliance/next-generation-compliance. 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 42.

The federal government is a significant generator of data.[117] 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 II.[118] 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 (NOAA) is using data generated by satellite technology to predict the future course of climate change.[119] The federal government uses a geographic information system (GIS) platform to help develop management policies for lands and resources administered by the Bureau of Land Management and the U.S. Forest Service.[120] As noted above, NOAA is also integrating geospatial data in GIS maps to serve as a resource for environmental emergency responders charged with dealing with events such as oil spills and natural disasters.[121]

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.[122] It has also created a mapping tool to assist in the identification of low-income and minority populations being subjected to disproportionate environmental burdens.[123]

Of direct relevance to environmental compliance and enforcement, EPA is relying on advanced technology in both the monitoring and reporting realms. An example of the use of the agency's use of advanced technology in the monitoring context to help identify regulatory violations involves infrared cameras. EPA has tested a computer program that relies on infrared pollution detection devices to measure emission rates that it expects will be useful in its enforcement efforts.[124] Infrared cameras allow users to detect the presence of compounds that are

---

[117] Madison, *supra* note 73, at 1611. The federal government also engages in the collection, aggregation, facilitation, and funding of data generation by non-governmental sources. *Id*. at 1612-20.

[118] *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).

[119] *See* Laurie J. Schmidt, *Twelve years of satellite data help decode climate change* (Apr. 14, 2015), http://climate.nasa.gov/news/2264/twelve-years-of-satellite-data-help-decode-climate-change/; National Atmospheric Administration, Global Climate Change: Vital Signs of the Planet, Taking a global perspective on Earth's climate, http://climate.nasa.gov/nasa_role/ (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-20th century are very likely due to increased atmospheric concentrations of greenhouse gases, and that NASA scientists and engineers intend to use data to answer questions such as the future course of temperatures and sea level rise).

[120] Breggin & Callan, *supra* note 35, at 131.

[121] *See supra* notes 81-82 and accompanying text; Breggin & Callan, *supra* note 35, at 136-37.

[122] Breggin & Callan, *supra* note 35, at 132.

[123] *Id*. at 132-33.

[124] Renee Schoof, *EPA Testing New Way to Measure Air Pollution Emissions*, 46 ENV'T REP. 3244 (2015) [hereinafter Schoof, *EPA Testing*]; *see also* Renee Schoof, *Infrared Camera Use Growing in Oil and Gas Sector*, 47 ENV'T REP. 1007 (2016) [hereinafter Schoof, *Infrared Camera*] (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

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.[125] 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.[126]

EPA has also modernized its Enforcement and Compliance History Online (ECHO) database, first established in 2002 to help communities assess environmental compliance, to better support frequent data updates and web services, enhance its interactive features, and improve display on mobile devices.[127] EPA enforcement officials characterize ECHO as "another potential resource to investors and communities."[128] They characterize the database as fostering better transparency, though the agency also acknowledges that a variety of data problems persist and that ECHO has not yet approached its potential value.[129] Over the years, while praising EPA for the effort, some commentators have characterized ECHO as a tool that has not yet achieved its full potential in increasing transparency.[130] More recently, the Government Accountability Office reported EPA's assertion that the public is making increasing use of the agency's ECHO website.[131] EPA's own Office of Inspector General concluded in 2016, however, that information obtained from the ECHO website pertaining to the regulatory 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.[132] According to the Inspector General, although 12 million ECHO

---

monitoring requirements). 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 42, at 112.

[125] Hindin & Silberman, *supra* note 42, at 112.

[126] *Id*. at 111-13.

[127] U.S. EPA, ECHO, Enforcement and Compliance History Online, What's New, https://echo.epa.gov/resources/general-info/whats-new. 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*, *Environmental Quality Management*, 12 ENVTL. QUALITY MGMT. 81 (Summer 2003).

[128] Hindin & Silberman, *supra* note 42, at 122.

[129] U.S. EPA, Enforcement and Compliance History Online, Known Data Problems, https://echo.epa.gov/resources/echo-data/known-data-problems (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 Dep't of Envtl. Prot., Maine Information relating to US EPA ECHO, http://www.maine.gov/dep/enforcement/echo.html (expressing concerns about ECHO's completeness and accuracy because of differences in vocabulary used by EPA and states, among other factors) [hereinafter Maine ECHO].

[130] *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.

[131] U.S. Gov't Accountability Off., *EPA Should Develop a Strategic Plan for Its New Compliance Initiative* 8, GAO-13-115 (2012), http://www.gao.gov/assets/660/650711.pdf.

[132] U.S. EPA, Off. of Inspector Gen., *Clean Air Act Facility Evaluations Are Conducted, but Inaccurate Data Hinder EPA Oversight and Public Awareness*, At a Glance, 9, Rep. No. 16-P-0164 (2016), https://www.epa.gov/sites/production/files/2016-05/documents/20160503-16-p-0164.pdf. The Inspector General

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."[133] Improvements to the quality and completeness of data in ECHO, achieved through e-reporting and other initiatives, is likely to enhance the value of the database over time.[134]

In addition, EPA will be responsible for engaging in follow-up investigatory and enforcement activities if data supplied by others (such the kinds of data generated by regulated entities and community groups, as described in the following sections) raise concerns about potential noncompliance, especially if the data supplied by those sources is deemed insufficient by itself to verify compliance status or support 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 non-governmental sources.

Finally, as explained in the next section, EPA has issued regulations requiring electronic reporting by regulated entities under statutes that include the Clean Water Act (CWA) and the Resource Conservation and Recovery Act (RCRA). Although the reporting of the data to EPA or the states is by the regulated entities themselves, EPA has created the infrastructure to facilitate the submission process. As EPA has explained, "E-reporting is not just 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 new e-reporting requirements in its 2015 NPDES e-reporting rule has the potential to have significant effects on EPA enforcement priorities for a variety of reasons, including because the rule's coverage of non-major facilities will give EPA access to information about compliance status of many regulated parties for the first time.[136]

B. The Use of Information Technology by Regulated Entities

---

found, for example, that of 65 facilities listed as major operating facilities, 26% were either closed, minor sources, never constructed, or not a facility. *Id*. at 9. EPA's response to the OIG report is included as Appendix B.

[133] *Id*. at 12.

[134] Markell & Glicksman, *Dynamic Regulation, Part II, supra* note 28.

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

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

> ("[T] the 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 nonmajor 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. NPDES Electronic Reporting Rule Implementation Guidance, 81 Fed. Reg. 62,395 (Sept. 9, 2016). For more thorough analysis of this potentially very significant development, see Markell & Glicksman, *Dynamic Regulation, Part II, supra* note 28.

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:

> 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]

---

[137] *See* F. William Brownell, *"Regulation by Guidance": A Response to EPA*, NAT. RESOURCES & ENV'T, Winter 1996, at 56, 57 (stating that in the early 1970s, EPA relied on stack tests "[b]ecause continuous emission monitors (CEMs) were unavailable").
[138] U.S. Sugar Corp. v. EPA, 2016 WL 4056404, *35 (D.C. Cir. 2016). *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, *Priority, 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. 10385, 10387 (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, *Hot Spots*, *supra* note 61, at 1478-79
[140] United States v. SVM Corp., 667 F. Supp. 1110, 1126 (D. Md. 1987).

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 monitoring that "measures emissions frequently to provide a representative measure of the monitored unit's continuous emission levels under applicable rules."[144] The agency successfully relied on continuous emissions monitoring 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. New technology is increasingly making CEM systems available "for a broad range of air emissions, including toxic substances, and water pollutants."[146] 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," although the agency 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

---

[141] U.S. Sugar Corp. v. EPA, 2016 WL 4056404, *38-41 (D.C. Cir. 2016) (noting that EPA decided to use an "upper prediction limit" (UPL), whose validity the court upheld, to establish regulatory standards, and that the UPL "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.").
[142] *Id*. at *54.
[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 using pollutant analyzer measurements and a conversion equation, graph, or computer program to produce results in units of the applicable emission limitation or standard." U.S. EPA, *Continuous Emission Monitoring – Information, Guidance, etc.*, https://www3.epa.gov/ttn/emc/cem.html.
[144] Hindin & Silberman, *supra* note 42, at 116; *see also* EPA, *Priority*, *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."). Continuous emissions monitoring "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. v. EPA, 2016 WL 4056404, *54 n.39 (D.C. Cir. 2016).
[145] *See* John Schakenbach, Robert Vollaro & Reynaldo Forte, *Fundamentals of Successful Monitoring, Reporting, and Verification under a Cap-and-Trade Program*, 56 J. OF THE AIR & WASTE MGMT. ASS'N 1576 (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, 4-8 (2010) (describing how continuous emissions monitoring equipment and automatic verification systems bolstered compliance levels under the CAA's acid rain program); Hindin & Silberman, *supra* note 42, 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] Hindin & Silberman, *supra* note 42, at 116.

facilities."[147] 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.[148] Others also see the potential for continuous emissions monitoring to provide greater reliability and greater credibility.[149] As CEM technology advances, so, too, will its reliability and value as a tool to enhance compliance and facilitate enforcement.

Another way to enhance the value of monitoring results for some purposes is to shift its location, rather than or in addition to its timing. EPA has 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 fence-line monitoring technologies and sensor network-based leak detection systems . . . ;
- Supplement air quality monitoring networks through development of low-cost, reliable air quality monitoring technology. . . ; and
- 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

[147] EPA, *Priority*, *supra* note 135, at 3.
[148] Hindin & Silberman, *supra* note 42, at 116.
[149] *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 & 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).
[150] Hindin & Silberman, *supra* note 42, at 116. EPA has noted the development of a new technology, differential absorption light detection, which can produce more accurate measurements of fugitive emissions from tanks. U.S. EPA, *Priority Next Generation Compliance Research Questions* 2 (May 18, 2016) (on file with authors). For further discussion of fenceline monitoring, see *supra* note 61 and accompanying text.
[151] U.S. 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 42, at 117.
[152] For a recent example of an enforcement action that includes advanced monitoring in new locations, see U.S. EPA, Tesoro and Par Clean Air Act Settlement, https://www.epa.gov/enforcement/tesoro-and-par-clean-air-act-settlement (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. . . ." U.S. EPA, Oil Refiners

refineries to deploy passive fenceline monitoring for benzene.[153] The rule requires regulated facilities to place monitors on the facility fenceline to measure facility emissions, specifies procedures for subtracting background concentrations and contributions to fenceline concentrations from other sources, and mandates corrective action if an applicable fenceline benzene concentration action level is exceeded.[154]

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, the traditionally available pollution monitoring technology, whose measurements tended to be sporadic and fixed at the emission point, created 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 pollution detection technology so regulated entities, the government, and the public can more easily 'see' pollutant discharges, environmental conditions, and

---

to Reduce Air Pollution at Six Refineries Under Settlement with EPA and Department of Justice (U.S. EPA Press Statement 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, 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 *Dynamic Governance, Part II*, *supra* note 28.

[153] Petroleum Refinery Sector Risk and Technology Review and New Source Performance Standards, 80 Fed. Reg. 75,177, 75,191-200 (Dec. 1, 2015); *id*. at 75,254-57 (codified at 40 C.F.R.§ 63.658).

[154] *Id*. at 75,192; *see also* Hindin & Silberman, *supra* note 42, at 116 (noting that fenceline monitoring can "serve as triggers for further monitoring or corrective actions by the facilities"). Regulations adopted under the Clean Water Act require whole effluent toxicity testing under certain conditions. 40 C.F.R. §§ 122.21(j)(5), 122.44(d)(1)(iv) (2015).

[155] Hindin & Silberman, *supra* note 42, at 116.

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 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 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.

EPA has already issued regulations requiring electronic reporting under the CWA's National Pollutant Discharge Elimination System permit program.[162] Pursuant to legislation adopted in 2012 to amend RCRA,[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

---

[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. *See* Renee Schoof, *Drone Use for Environmental Monitoring May Grow under Rule*, 47 ENV'T REP. (BNA) 2609, 2609 (2016) (quoting a remote sensing specialist).

[157] Hindin & Silberman, *supra* note 42, at 118.

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

[159] EPA, *Priority*, *supra* note 135, at 9.

[160] Hindin & Silberman, *supra* note 42, 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, *Priority*, *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).

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

[164] U.S. EPA, Hazardous Waste Electronic Manifest System ("e-Manifest"), https://www.epa.gov/hwgenerators/hazardous-waste-electronic-manifest-system-e-manifest/; Hazardous Waste

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] EPA has acknowledged as much.[168] Differences in EPA and state vocabulary, among other factors, suggest the need for additional work to improve accuracy with respect to both government actions and regulated party performance.[169]

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. 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]

Another technology that EPA regards as a promising compliance enhancement tool is immediate feedback technology. 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]

Manifest System; Modification of the Hazardous Waste Manifest System; Electronic Manifests, 79 Fed. Reg. 7517 (Feb. 7, 2014).

[165] Hindin & Silberman, *supra* note 42, at 118.

[166] EPA, *Priority*, *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 misreporting of renewable fuel credits under a regulatory program implemented under the Energy Policy Act of 2005).

[168] Giles, *supra* note 19, at 26 ("And where government relies on self-reporting for compliance data, we also need ways to check for accuracy.").

[169] *See, e.g.*, Maine ECHO, *supra* note 129 (expressing concerns about ECHO's completeness and accuracy because of differences in vocabulary used by EPA and states, among other factors).

[170] Hindin & Silberman, *supra* note 42, at 118 n.137.

[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 (HAPs), Professor McGarity concluded nearly a decade ago that:

These officials hope that these mechanisms will yield "positive behavioral impacts" because of their ability to create in regulated entities a perception of increased risk of detection of noncompliance, a "classic deterrence response."[173] The potential obstacles to use of the technology as a compliance promoter include unresolved questions concerning its cost-effectiveness and EPA's legal authority to require its use.[174] If those questions yield positive answers, EPA sees the value of requiring its use through regulations or enforcement settlements.[175]

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.[176] As the discussion below indicates, individuals, citizen groups, and communities now have access to affordable sensors and other monitoring devices capable of generating data on both ambient environmental conditions and regulatory compliance.[177] 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."[178]

> 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 HAPs. 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, *Hot Spots*, *supra* note 61, 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 42, at 111.

[174] *Id*.

[175] *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*.

[176] Terry, *supra* note 37, 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 27, at 10987-90; *see also* ELI, BIG DATA, *supra* note 37, at 3-29 (also including data generation efforts by non-governmental organizations).

[177] *See infra* notes 184-188 and accompanying text.

[178] ELI, BIG DATA, *supra* note 37, at 23. Although this section 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 28, at Parts III and IV; *Dynamic Governance, Part II*, *supra* note 28, at Parts IIC and III.

Some refer to the phenomenon as "citizen science."[179] 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, "to accelerate the use of crowdsourcing[180] 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."[181] The site's aim is to help federal agencies "accelerate innovation through public participation."[182] A non-government 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 which "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."[183]

One reason that data generated by non-governmental sources has begun to proliferate is the increased access to the technology that generates the data.[184] Access has increased because the

---

[179] 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." ELI, CLEARING THE PATH, *supra* note 69, at 5. For a more in depth review of the role of citizens in environmental enforcement, see *Dynamic Governance, Part I*, *supra* note 28, at Part IV.

[180] 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&.

[181] citizenscience.gov, *Helping Agencies accelerate innovation through public participation*, http://www.citizenscience.gov/ [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 16 federal and state agencies that related to ecology and environment, however, and one of those was EPA's Urban Waters Program-Amigos Bravos. https://ccsinventory.wilsoncenter.org/#projectId/132. For discussion of the role of citizen science in public policymaking, see generally ELI, CLEARING THE PATH, *supra* note 69.

[182] citizenscience.gov, *supra* note 181.

[183] Brian L. Sullivan et al., *The eBird Enterprise: An Integrated Approach to Development and Application of Citizen Science*, 169 BIOLOGICAL CONSERVATION 31 (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.

[184] *See* Fahey, *supra* note 37, 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").

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."[185] The devices that can generate data of potential value in the implementation of government programs 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.[186] 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.[187] The increased access to this array of technology has led to what has been called the "democratization" of the collection of data of potential value to the implementation of regulatory programs.[188]

2. *The Role of Citizen Science in Environmental Enforcement*

Government agencies, including both federal and state environmental agencies, have welcomed the development and increased availability of these technologies and taken steps to engage citizens in these kinds of efforts. EPA Administrator Gina McCarthy, for example, has touted the use of smart phones as air quality monitors.[189] Community-driven generation of

---

[185] Madison, *supra* note 73, at 1610; *see also* Fahey, *supra* note 37, at 330 ("[T]he major news item regarding the democratization of big data is that 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 27, at 10985 (citing Emily G. Snyder et al., *The Changing Paradigm of Air Pollution Monitoring*, 47 ENVTL. SCI. & TECH. 11369 (2013)); Rodriguez, *supra* note 80 (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 66; *see also* Ehrich, *supra* note 152 ("[A]dvances in information and monitoring technologies increasingly put portable, lower-cost monitoring devices into the hands of individuals or groups interested in air or water quality in their personal environments.").

[186] Batty, *supra* note 64, 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/.

[187] Hindin & Silberman, *supra* note 42, at 112.

[188] Fahey, *supra* note 37, 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 57, 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 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).

[189] Pat Rizzuto, *Get Ready for Phone Air Monitor Data, EPA's McCarthy Says*, 47 ENV'T REP. 1029 (2016). At least one non-environmental federal agency has actually created a smart phone app to solicit the public's help in

information on environmental conditions has begun contributing to federal and state government efforts across the nation to implement and enforce the environmental laws. Various citizen initiatives, which are not limited to but include those designed to promote compliance with regulatory obligations, were either encouraged by government officials or provided information of potential value to them. Examples include the Virginia Department of Environmental Quality's establishment of different levels of required data quality depending on the intended use of citizen-science monitoring data,[190] the use of infrared cameras by Colorado workers to detect methane leaks in natural gas wells,[191] detection of fugitive emissions from sources in Boston,[192] the use by university researchers of vans to identify the sources of benzene emissions in Houston neighborhoods,[193] water quality sampling along the Mississippi River basin by volunteers organized by the University of Nebraska-Omaha in an effort to detect and determine the sources of atrazine pollution,[194] the use of an iPhone application that IBM developed to make it possible for citizens to monitor water quality,[195] and the Freshwater Trust's use of boats equipped with cameras linked to Google Maps to assess surface water quality in Oregon.[196] 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.[197]

---

identifying violators of the law. *See* U.S. Immigration and Customs Enforcement, *Operation Predator App*, https://www.ice.gov/predator/smartphone-app (for use in apprehending child predators).

[190] *See id*. at 31.

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

[192] *See id*.

[193] 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), http://www.houstonpress.com/news/the-way-we-currently-monitor-air-pollution-near-the-ship-channel-sucks-researchers-say-8263528.

[194] Rejeski & McElfish, *supra* note 70, at 64.

[195] Hindin & Silberman, *supra* note 42, at 113 n.93.

[196] Frederick Reimers, *Mapping America's Disgusting Waterways*, BLOOMBERG BUSINESSWEEK, Oct. 15, 2015, 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, *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, http://globalfishingwatch.org/. In addition, with funding by the Bureau of Land Management, volunteers have helped track sage grouse populations on public lands. Rejeski & McElfish, *supra* note 70, at 64. *See also supra* note 183 and accompanying text (discussing the eBird enterprise for collecting data on bird distribution and abundance).

[197] For example, if monitoring of ambient conditions in Houston to determine whether neighborhoods are being exposed to unsafe levels of benzene emissions identify the likely sources contributing to excessive ambient benzene

Because quality control over citizen science is not likely to match an environmental agency's own information-gathering efforts, agencies have used data supplied by non-governmental sources as a signal that further inquiry by the agency into compliance status or ambient conditions is warranted. 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.[198] 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 VW case itself.[199]

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 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.[200] 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.[201]

If agencies such as EPA are to take advantage of the new streams of data being generated and flowing to them, they will need to develop protocols for collecting, storing, processing, and using the information. EPA has 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.[202] 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.[203] The agency has posted its Air Sensor Toolbox for Citizen Scientists, which provides guidance on sampling methodologies, calibration and validation approaches, measurement

---

concentrations, that information may spur further investigation by federal or state regulators to determine if those sources are violating permit limits.

198 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.

199 Evidence of irregularities in VW 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.

200 *See* Rejeski & McElfish, *supra* note 70, at 63.

201 *See* ELI, CLEARING THE PATH, *supra* note 69, at 25.

202 Stuart Parker, *EPA Takes Early Steps to Bolster Agency's Future Use of 'Citizen Science'*, INSIDE EPA, ENVTL. POLICY ALERT, May 11, 2016, at 25.

203 Yardley, *supra* note 191.

methods options, data interpretation, and low-cost sensor performance.[204] The site includes links to videos on EPA-hosted training workshops,[205] descriptions of EPA's research on air sensor monitoring and analysis technologies,[206] information about a project designed to demonstrate the capabilities of new real-time monitoring technology (called the Village Green Project),[207] and a series of links relating to a free, web-based tool (called Real-Time Geospatial Data Viewer) that can show air quality data collected by individuals while walking, biking, or driving.[208]

Similar developments are occurring at the state level. As indicated above,[209] Virginia environmental officials developed criteria for the appropriate use of data provided by non-governmental monitors, including the identification of waters for follow-up monitoring by the state agency.[210] 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.[211]

D. Information Technology Advances and Enhanced Transparency

An additional benefit of the new information technology that is available to agencies and non-governmental entities is the potential to enhance transparency, a principal goal not only of Next Gen but of executive branch policy more generally.[212] The federal government maintains, aggregates, and makes available to the public a huge array of data concerning myriad aspects of society, including many categories that are relevant to environmental law and policy, such as ecosystems, climate, public safety, and energy.[213] As some of the examples above demonstrate,

---

[204] U.S. EPA, *Air Sensor Toolbox for Citizen Scientists*, https://www.epa.gov/air-research/air-sensor-toolbox-citizen-scientists.
[205] U.S. EPA, *Community Air Monitoring Training*, https://www.epa.gov/air-research/community-air-monitoring-training.
[206] U.S. EPA, *Air Monitoring, Measuring, and Emissions Research*, https://www.epa.gov/air-research/air-monitoring-measuring-and-emissions-research#ngamt.
[207] U.S. EPA, *Village Green Project*, https://www.epa.gov/air-research/village-green-project.
[208] U.S. EPA, *Real Time Geospatial Data Viewer (RETIGO)*, https://www.epa.gov/hesc/real-time-geospatial-data-viewer-retigo.
[209] *See supra* note 69 and accompanying text.
[210] *Citizen Water Quality Data*, *supra* note 70; VA. DEP'T OF ENVTL. QUALITY, CITIZEN MONITORING GUIDANCE, http://www.deq.virginia.gov/Programs/Water/WaterQualityInformationTMDLs/WaterQualityMonitoring/CitizenMonitoring/Guidance.aspx; *see also* Hindin & Silberman, *supra* note 42, at 121 (describing Virginia's citizen water quality monitoring program as an effective example of citizen participation in environmental monitoring).
[211] Water Action Volunteers, *Water Action Volunteers Stream Monitoring Program*, http://watermonitoring.uwex.edu/wav/. The Riverkeeper network also engages volunteers to help monitor surface water quality. *See, e.g.*, Potomac Riverkeeper Network, *What Is a Riverkeeper?*, http://www.potomacriverkeepernetwork.org/riverkeepers/.
[212] *See. e.g.*, Exec. Order No. 13642, Making Open and Machine Readable the New Default for Government Information, 78 Fed. Reg. 28111 (May 14, 2013). Increased transparency has been a long-standing goal of the federal government, and of EPA. *See, e.g.*, David Markell, *"Slack" in the Administrative State and its Implications for Governance: the Issue of Accountability*, 84 OR. L. REV. 1 (2005) [hereinafter Markell, *Slack*].
[213] Data.gov, https://www.data.gov/.

EPA does not intend to keep the fruits of data supplied by individuals and communities to itself. Instead, the emergence of technological innovations in the fields of monitoring and reporting have the potential to transform the agency's capacity to serve as a clearinghouse of information, one of the long-standing functions of the agency.[214] ECHO, summarized above, is a prominent example of EPA's efforts to serve in this capacity in the compliance realm.[215]

Newly-available and more easily accessible information is likely to have an enormous variety of effects on the roles of government, regulated parties, and citizens. According to EPA enforcement officials, expanded transparency with respect to new information stemming from advanced emissions monitoring and e-reporting can "provide more accurate, complete, and timely information on pollution sources, pollution, and compliance," empowering "communities and the marketplace to play a more active role in compliance oversight and improve the performance of both the government and regulated entities."[216] 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.[217] That information is likely to help inform priority-setting by government officials. It is also likely to make enforcement easier and more likely, in at least some cases, including by citizens. Such a rise in citizen enforcement may increase the need for coordination of enforcement efforts.[218] Information about relative performance also has the potential to put pressure on lower performing companies to improve their performance.[219]

The increased information about noncompliance may also increase the appropriateness of enforcement action to address violations. EPA's transformation of its compliance and enforcement program may not be due entirely to technological developments and perceived failings in existing practices. In particular, EPA's interest in taking advantage of citizen science may be due in part to resource concerns.[220] Resource constraints have made it difficult to sustain the kind of

---

[214] Hannah J. Wiseman, *Regulatory Islands*, 89 N.Y.U. L. REV. 1661 (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, *Slack*, *supra* note 212, 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 (1993) (discussing the federal government's potential to serve as a clearinghouse).

[215] *See supra* notes 127-134 and accompanying text. We do not mean to understate the extent of the challenge in obtaining and managing data. *See supra* note 214.

[216] Hindin & Silberman, *supra* note 42, at 103.

[217] *See* Giles, *supra* note 19, at 26.

[218] *See* David Freeman Engstrom, *Agencies as Litigation Gatekeepers*, 123 YALE L.J. 616 (2013) (discussing role of agencies as "gatekeepers" of private enforcement activity).

[219] Giles, *supra* note 19, at 26.

[220] 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 42, at 103.

enforcement presence the agency has traditionally had.[221] Newly-available information about noncompliance made possible by the introduction of e-reporting and other developments that enhance information about compliance status is likely to compound such challenges.[222] The government will need to be mindful of limits in citizen capacity,[223] and of maintaining appropriate levels of enforcement capacity internally.

A third consequence of increased and more easily accessible data involves its potential to spur follow-up efforts to learn more about risks and their possible sources. Information about unacceptable ambient conditions may spur those situated upwind or upstream from the problem to engage in follow-up monitoring to determine the responsible source or 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.[224] In some cases, new data, or newly-organized data may reveal upstream regulatory gaps that motivate the agency and others to plug. EPA has used data compiled in the Toxic Release Inventory compiled under the Emergency Planning and Community Right-to-Know Act,[225] for example, to not only monitor compliance with existing regulatory standards and identify enforcement priorities, but also "to help assess whether new regulations are needed to address environmental problems."[226]

A fourth benefit of citizen-generated data, in particular, at least in the view of some commentators, is its potential to "generate data where formal institutions have failed to do so, empower the public by giving them ownership over the data collection process, and teach data literacy through the act of collection."[227]

---

[221] *See, e.g.*, Joel A. Mintz, *"Running on Fumes": The Development of New EPA Regulations in an Era of Scarcity*, 46 ENVTL. L. REP. 10510 (2016).
[222] *See Dynamic Governance*, *Part II*, *supra* note 28 (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).
[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, 39–40 (2014) (discussing a case study in which even citizens in an affluent community were overwhelmed by governance duties).
[224] *See* Bass, *supra* note 70, 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 198-199 and accompanying text (discussing EPA's follow-up investigations concerning Volkswagen's defeat devices); Macey, *supra* note 56, 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").
[225] *See* 42 U.S.C. § 11023 (2012); U.S. EPA, *Toxic Release Inventory (TRI) Program*, https://www.epa.gov/toxics-release-inventory-tri-program.
[226] John D. Echeverria & Julie B. Kaplan, *Poisonous Procedural "Reform": In Defense of Environmental Right-to-Know*, 12-SPG. KAN. J.L. & PUB. POL'Y 579, 583 (2003).
[227] Williams, *supra* note 110, at 8.

A final example involves the 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 (e.g., between regulated parties and community groups, among regulated parties, or among community groups).[228]

Critically, as EPA has acknowledged, information must be correct for transparency to work effectively in these ways.[229] 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.[230] As we have suggested above,[231] work remains to be done in this arena, but progress is being made.

IV. CONCLUSION

Some have argued that the construction of the federal environmental statutes and the realities of modern technology are fundamentally mismatched.[232] Gregg 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," and that "the conversion of data into useful, policy-relevant knowledge will change dramatically," replacing the 'architecture of ignorance' that is currently in place."[233]

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 improved monitoring capacity, advances in information dissemination

---

[228] Giles, *supra* note 19, at 24 (noting that new data may facilitate interactions between regulated parties and nearby communities).
[229] *Id*. at 26; Katrina Fischer Kuh & David L. Markell, *Informational Regulation, the Environment, and the Public* (forthcoming 2016) (reviewing several informational regulation strategies and identifying accuracy of information as an important element); *see also supra* Part IIB.1 (discussing data quality challenges).
[230] 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.").
[231] *See supra* Part IIB.1-2.
[232] *See, e.g.*, Macey, *supra* note 57, 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 29, 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.").
[233] Macey, *supra* note 57, at 1631; *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.").

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." 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 pose, the prospects for success of its effort to transform its enforcement and compliance programs are likely to 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 is reflected in a myriad of examples. These include instances in which community groups and environmental activists generate information for EPA, and 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 thereby improve compliance with regulatory requirements.