



2017

Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency

Beth Simone Noveck

Tandon School of Engineering, New York University

Follow this and additional works at: <https://digitalcommons.law.yale.edu/yhrdlj>



Part of the [Human Rights Law Commons](#)

Recommended Citation

Beth Simone Noveck, *Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency*, 19 *YALE HUM. RTS. & DEV. L.J.* (2017).

Available at: <https://digitalcommons.law.yale.edu/yhrdlj/vol19/iss1/1>

This Article is brought to you for free and open access by Yale Law School Legal Scholarship Repository. It has been accepted for inclusion in Yale Human Rights and Development Law Journal by an authorized editor of Yale Law School Legal Scholarship Repository. For more information, please contact julian.aiken@yale.edu.

Essay

Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency

Beth Simone Noveck[†]

Open data policy mandates that government proactively publish its data online for the public to reuse. It is a radically different approach to transparency than traditional right-to-know strategies as embodied in Freedom of Information Act (FOIA) legislation in that it involves ex ante rather than ex post disclosure of whole datasets. Although both open data and FOIA deal with information sharing, the normative essence of open data is participation rather than litigation. By fostering public engagement, open data shifts the relationship between state and citizen from a monitorial to a collaborative one, centered around using information to solve problems together. This Essay explores the theory and practice of open data in comparison to FOIA and highlights its uses as a tool for advancing human rights, saving lives, and strengthening democracy. Although open data undoubtedly builds upon the fifty-year legal tradition of the right to know about the workings of one's government, open data does more than advance government accountability. Rather, it is a distinctly twenty-first century governing practice borne out of the potential of big data to help solve society's biggest problems. Thus, this Essay charts a thoughtful path toward a twenty-first century transparency regime that takes advantage of and blends the

[†] Beth Simone Noveck is the Jerry M. Hultin Global Network Professor at the Tandon School of Engineering, New York University, where she directs the Governance Lab. Formerly the director of the White House Open Government Initiative, she is the author of *Smart Citizens, Smarter State: The Technologies of Expertise and the Future of Governing* (Harvard University Press 2015) and *Wiki Government: How Technology Can Make Government Better, Democracy Stronger, and Citizens More Powerful* (Brookings 2010).

strengths of open data’s collaborative and innovation-centric approach and the adversarial and monitorial tactics of freedom of information regimes.

TABLE OF CONTENTS

INTRODUCTION.....3

I. HOW OPEN DATA DIFFERS FROM FREEDOM OF INFORMATION POLICY6

 A. Differences in Timing.....7

 B. Differences in Information Types.....8

 C. Differences in Audience.....9

II. THE OPEN DATA MOVEMENT: THE FUEL FOR DATA-DRIVEN ACTIVISM.....10

 A. History of U.S. Federal Initiatives in Open Data.....11

 B. Other Governments, Corporations, Civil Society: The Open Data Movement Takes Hold14

III. FROM DATA TO ACTION: OPEN DATA’S IMPACTS18

 A. Analytical Opportunities.....18

Identifying Inequities and Inefficiencies by Measuring Past Performance18

Targeting Allocation of Scarce Resources18

Predicting Outcomes.....19

 B. The Social Impacts of Open Data.....19

Improving Governmental Accountability.....19

Improving Accountability of Private Actors.....20

Enhancing Consumer Choice Through Smart Disclosure21

Promoting Entrepreneurship22

Saving Lives and Solving Problems.....23

IV. THE MORAL DATA ECONOMY: OPEN DATA, HUMAN RIGHTS, AND DEVELOPMENT.....24

 A. Facilitating Empirical Social Science, Investigative Journalism, and Consumer Protection.....24

 B. Advancing Civil Rights and Other Public Interest Litigation by Revealing Disparate Impacts25

 C. Assessing the Effectiveness and Fairness of Institutions of Justice26

 D. Uncovering Human Rights Abuses.....28

 E. Reducing Abuse and Enhancing the Impact of Development Assistance.....31

V. OPEN DATA AND FOIA: COMPLEMENT NOT REPLACEMENT32

 A. Open Data’s Potential Shortcomings32

Challenge 1: Political Will32

Challenge 2: The Lack of a Right of Action34

Challenge 3: Data Invisibles.....36

<i>Challenge 4: Prioritization</i>	37
B. Proposals for Enhancing Both FOIA and Open Data	39
<i>Proposal 1: Merge Open Data with FOIA Collection and Publication Mechanisms</i>	40
<i>Proposal 2: Publish All Information Requested Pursuant to Either Freedom of Information or Open Data Law in Machine-Readable Formats</i>	41
<i>Proposal 3: Increase Dialogue Between Officials Responsible for FOIA and Those Responsible for Open Data</i>	41
<i>Proposal 4: Create and Store Data Digitally in a Searchable Cloud</i>	43
VI. CONCLUSION	44

INTRODUCTION

Created by the Mexico Institute of Competitiveness (IMCO), *Mejora Tu Escuela*¹ is an online platform that makes government data about Mexico's schools publicly available. The website provides parents with comparative data so that they can compare their own school's results to others, thereby empowering them to demand better-quality education for their children. It publishes expenditure data, giving activists, administrators, policymakers, and journalists the means to dig deeper, to spot fraud and corruption, and to advocate for change. This is exactly what happened in 2014, when a report by IMCO revealed that over 1,400 teachers on public school payrolls were supposedly older than a hundred years old (with most having the same birthday)² and that many earned more than the president of Mexico.³

No, the school board had not discovered Ponce de León's Fountain of Youth. Rather, the story of *Mejora Tu Escuela* illustrates how, when government makes information free of charge and readily downloadable in digital form, such open data can inform citizen engagement and activism. With the availability of information in machine-readable formats, those with technical knowhow, whether they are the data owners or not, can create tools, models, and analyses that enable empirical insights and data-driven solutions to societal problems.⁴

In this case, federal authorities had required states to provide information about the condition of schools, payrolls, and other expenditures.⁵ But it was civil society activists at IMCO who created the

1. MEJORA TU ESCUELA, <http://mejoratuescuela.org> [<https://perma.cc/M5T7-3J6K>].

2. STEFAAN VERHULST & ANDREW YOUNG, *THE GLOBAL IMPACT OF OPEN DATA: KEY FINDINGS FROM DETAILED CASE STUDIES AROUND THE WORLD 207* (2016) (ebook).

3. *Id.*

4. See *Open Data: What's In a Name?*, THE GOVERNANCE LAB [GOVLAB] (Jan. 16, 2014), <http://thegovlab.org/open-data-whats-in-a-name> [<https://perma.cc/WCV5-SDJ3>].

5. VERHULST & YOUNG, *supra* note 2, at 198.

platform to make that information accessible to citizens and who also scrutinized that information,⁶ ultimately exposing rampant malfeasance that was previously hidden. Although the government initially prevaricated, claiming clerical error, the ensuing media frenzy over the website helped to prompt reform and a shift of responsibility over education from states to the federal government. Ultimately, the activists and the federal bureaucracy worked in parallel, addressing this local-level corruption and acting to improve Mexico's schools.⁷

By fostering greater public engagement and collaboration, open data represents a major governing innovation in the twenty-first century and a potentially important tool for advancing human rights and saving lives. Although open data undoubtedly builds upon and grows out of a fifty-year legal tradition of freedom of information and the right to know about the workings of one's government in a democracy, open data does more than advance government transparency. Rather, it is a distinctly twenty-first century approach borne out of the potential of big data to help solve society's biggest problems.

Although Freedom of Information Act (FOIA) requestors are also those clamoring for government information in order to solve a problem, open data works differently than traditional right-to-know strategies as embodied in FOIA legislation. FOIA is an inherently adversarial tactic focused on prying secrets out of government. Open data is not. Openly publishing the information instead attracts collaboration by knowledgeable and passionate members of the public, who augment the manpower and skills often lacking in under-resourced public institutions.

Because governments in an open data regime must proactively publish their data with the intent that people use it, the normative essence of open data is participation rather than litigation. By catalyzing public engagement—both scrutiny of data by the public and collaboration with the public in building new analytical tools and websites—open data is, in and of itself, profoundly shifting the political economy of transparency and changing the relationship between the public and the state vis-à-vis information. If, as David Pozen suggests, FOIA is deeply reactionary and advances neo-liberal ends by tying up the machinery of the regulatory state,⁸ by contrast, the open data legal framework may be advancing participatory democracy.

Whereas FOIA promotes deliberative discourse about what government *did*, open data anticipates what institutions and citizens *can do together* to create value of different kinds, especially to advance evidence-based policymaking and co-creation of solutions to hard problems. Whereas FOIA describes a process for identifying *what* data to publish, open data describes a set of technology standards that determine *how* to

6. *Id.* at 201-02.

7. See TIM KELSEY & ROGER TAYLOR, *TRANSPARENCY AND THE OPEN SOCIETY: PRACTICAL LESSONS FOR EFFECTIVE POLICY* 146-151 (2016).

8. David E. Pozen, *Freedom of Information Beyond the Freedom of Information Act*, 165 U. PA. L. REV. 1097, 1101 (2017).

publish information.⁹ Whereas FOIA is *rights-based*, open data is *technology-driven*. As a result, while freedom of information activists usually come from the good government or human rights communities, open data activists, at least in these early days of the movement, have often been technology-savvy “civic hackers” with a different set of skills and approaches.¹⁰

By shifting the underlying theoretical understanding of the relationship between the state and the public from the adversarial to the collaborative, open data both complements and complicates our reliance on FOIA as the bedrock of the public’s right to know about the workings of government. FOIA is rooted in a narrative about hard-won transparency increasing governmental accountability, but this has not always borne itself out in practice,¹¹ as FOIA has been fraught with problems and limitations. Delays in responses and redactions frustrate information seekers, while the volume of requests, including politically-motivated nuisance requests, bedevils government agencies.¹²

In addition, since the era of modern FOIA regimes began with the United States’ adoption of the Freedom of Information Act in 1966,¹³ rates of trust in government around the world have steadily declined to all-time lows.¹⁴ This may not be coincidental: because FOIA is used in adversarial situations, FOIA tends to highlight the worst of government by demonstrating how public officials have tried to hide misdeeds. It emphasizes malfeasance, invariably shaping public perception of government. In addition, the bulk of FOIA requests in the United States have come primarily from corporations using them as tools to advance their business interests rather than from investigative journalists or NGOs.¹⁵ Potentially “flawed beyond repair,”¹⁶ as Pozen writes, FOIA may foster

9. The International Open Data Charter represents a notable and widely accepted example of such guidelines. See OPEN DATA CHARTER, <http://opendatacharter.net> [<https://perma.cc/G2HM-8AFT>]; *Adopted By*, OPEN DATA CHARTER, <http://opendatacharter.net/adopted-by-countries-and-cities> [<https://perma.cc/43B2-4W5B>]; *Endorsed By*, OPEN DATA CHARTER, <http://opendatacharter.net/endorsed-by> [<https://perma.cc/2YVG-5C8G>].

10. Katleen Janssen, *Open Government Data and the Right to Information: Opportunities and Obstacles*, 8 J. COMM. INFORMATICS (2012), <http://ci-journal.net/index.php/ciej/article/view/952/954> [<https://perma.cc/3FSE-7PPN>].

11. Pozen, *supra* note 8, at 1129 n.190 (“It is clear (and not at all speculative) that the work of many federal officials has been publicized through FOIA, which generates retail accountability in the form of reputational and other consequences when that work is seen to be faulty. What is far less clear is whether FOIA has improved the quality of administrative action in a larger, lasting sense.”).

12. *Id.* at 1123.

13. Ben Worthy, *More Open But Not More Trusted? The Effect of the Freedom of Information Act 2000 on the United Kingdom Central Government*, 23 GOVERNANCE 561, 561 (2010).

14. Americans’ trust in government continues to hover near its historic lows. See *Public Trust in Government: 1958-2017*, PEW RES. CTR. (May 3, 2017), <http://www.people-press.org/2017/05/03/public-trust-in-government-1958-2017> [<https://perma.cc/DP65-X3U6>]. For discussion of this issue in other nations, see Worthy, *supra* note 13, at 574-577.

15. Pozen, *supra* note 8, at 1103.

16. *Id.* at 1136.

litigation without better government to show as a result.

Open data and FOIA bear many similarities, but open data places at the center not the right to know but rather *innovation* – in addressing public challenges and engaging with citizens. To be clear, open data will not be a panacea for all social challenges, and not all government information is or should be shared openly. Furthermore, and most critically, open data relies on the willingness of the data owner to publish the data, as open data policies generally do not have the “teeth” to compel disclosure where the information holders may be reluctant to do so. In such situations, freedom of information law remains necessary to compel disclosure, at least until other enforcement mechanisms can be put in place. In that way, freedom of information law provides the necessary and complementary mechanism for citizens seeking information that has not yet been made publically available in the needed form.

The goal of this Essay is to chart a thoughtful path toward a twenty-first century transparency regime that takes advantage of the strengths of both open data’s collaborative and innovation-centric approach and the adversarial and monitorial tactics of freedom of information regimes. First, I outline the ways open data and older approaches differ. Although primarily discussing freedom of information through the lens of the U.S. Freedom of Information Act¹⁷ (and using FOIA as shorthand for freedom of information laws, generally), my analysis equally applies to right to information practices in the more than one hundred countries which have adopted them in some form.¹⁸ Second, I track the evolution of the open data movement and examine the hallmarks of open data policy and legislation around the world. Third, I focus specifically on the potential impact of open data as a tool for advancing the causes of economic development, human rights, and social justice. Fourth, I look at the challenges and weaknesses of FOIA and open data and offer recommendations on how to blend the best of both approaches to promote evidence-based, accountable governance. Finally, I conclude by suggesting that, in the long term, FOIA and open data may themselves converge as we move to a future where all government data sits in a secure but readily-accessible cloud, enabling both targeted provision of government services and far-reaching citizen engagement and collaboration.

I. HOW OPEN DATA DIFFERS FROM FREEDOM OF INFORMATION POLICY

Since its enactment in 1966, the Freedom of Information Act has served as the primary tool for citizens seeking government disclosure.¹⁹ Rooted in the conception that the public has a “right to know” about government’s

17. 5 U.S.C. § 552 (2012).

18. Freedominfo.org lists 115 countries with FOI regimes. See *Chronological and Alphabetical Lists of Countries with FOI Regimes*, FREEDOMINFO.ORG (June 30, 2016), <http://www.freedominfo.org/?p=18223> [<https://perma.cc/4K5N-V6LN>].

19. See WENDY GINSBERG, CONG. RESEARCH SERV., R41933, THE FREEDOM OF INFORMATION ACT (FOIA): BACKGROUND, LEGISLATION, AND POLICY ISSUES 2 (2014).

activities and operations, FOIA establishes a presumption of public accessibility of information held by the executive branch.²⁰ Under the law, any person can request agency records on any topic not falling within one of nine categories of exemptions.²¹

Open data clearly differs from FOIA in three ways: open data changes the timing for disclosure, focuses on different types of information, and addresses a broader audience.

A. Differences in Timing

Open data significantly shifts the default time of disclosure. Whereas FOIA institutionalizes *ex post* disclosure pursuant to a specific demand by individual requestors, open data thrives on *ex ante*, proactive publication of whole classes of information publicly and online, often in a centralized repository such as Data.gov.²² Under the Office of Management and Budget's (OMB) 2009 Open Government Directive, federal agencies are required to identify "high-value information" not yet available and to establish timelines for publication of those datasets.²³ Although high-value information might be released in response to a public consultation, it could just as likely be any information that "can be used to increase agency accountability and responsiveness; improve public knowledge of the agency and its operations; further the core mission of the agency; [and] create economic opportunity."²⁴ For example, under the Texas open data law,²⁵ which imitates the federal regime, the State Attorney General has proactively made key criminal justice datasets available for download, such as the list of Parole Absconders, Attorney General Opinions, Child Support Evaders, and Custodial Death Reports that might then aid in empirical criminal justice reform.²⁶ Taiwan publishes the country's air quality, water quality, and land ownership data.²⁷ High quality data results in the publication of many unexpected treasure troves of information collected by government such as the United Kingdom's Annual Statement on the Government Wine Cellar²⁸ or the Indian government's state and district

20. *See id.* at 1.

21. *Id.* at 3-4.

22. DATA.GOV, <https://www.data.gov> [<https://perma.cc/9ES6-SVB3>].

23. Peter R. Orszag (Director, Office of Mgmt. & Budget), *Memorandum for the Heads of Executive Departments and Agencies on the Open Government Directive*, EXEC. OFFICE OF THE PRESIDENT 7 (Dec. 8, 2009), https://obamawhitehouse.archives.gov/sites/default/files/omb/assets/memoranda_2010/m10-06.pdf [<https://perma.cc/3KE6-3L9Z>].

24. *Id.* at 7-8.

25. 2011 Tex. Gen. Laws 3850.

26. *High-Value Data Sets*, OFFICE OF THE ATT'Y GENERAL OF TEXAS (KEN PAXTON), <https://www.texasattorneygeneral.gov/og/high-value-data-sets> [<https://perma.cc/HL8L-6GPQ>].

27. *Global Open Data Index: Taiwan*, OPEN KNOWLEDGE INT'L, <https://index.okfn.org/place/tw> [<https://perma.cc/GJW5-G4BE>].

28. *Annual Statement on the Government Wine Cellar for the Financial Year 2014 to 2015*, GOV.UK (July 16, 2015), <https://www.gov.uk/government/publications/annual-statement-on-the-government-wine-cellar-for-the-financial-year-2014-to-2015> [<https://perma.cc/8P6F->

livestock census of donkeys, rabbits, dogs, and elephants.²⁹

B. Differences in Information Types

Second, open data emphasizes disclosure of different classes of information than does FOIA. FOIA law focuses more narrowly on accountability and, by extension, information on how government works. This data *created* by the government about the workings of government is what Cass Sunstein refers to as “input transparency.”³⁰ Open data goes further than FOIA, where requests paradigmatically focus on data about government operations. Open data provisions also strive to provide data *collected* by the government about the economy, environment, and society in its capacity as a scientific research organization or regulator.

Transparency in the open data context goes beyond the deliberations of government, the schedules of parliamentarians and ministers, the spending of treasuries, and the data that government produces about its own operations. “High-value data” includes an ever-expanding universe of datasets.³¹ These datasets include information gathered through governmental oversight and evaluation (such as workplace safety and injury records, airplane flight logs, and doctors’ prescriptions), as well as information gathered in governmental capacities as scientific research organizations (such as weather data and information about the human genome).

This expansion of scope is not uncontroversial. Critics fear it distracts from the need to publish certain essential datasets about the workings of government. They worry that “high-value data” suffers from inherent moral relativism and ambiguity (high-value to whom?).³² Publishing data about the locations of public toilets and park benches or data to spur entrepreneurship, for example, may allow public bodies to shirk their transparency obligations too easily. Yet this broader definition has also ushered in a movement toward more proactive disclosure of more kinds of information across all levels of government and an increasing realization that one of the reasons to release data openly is that “you never know who may derive benefit from it, and the ‘long tail of data’ may yield plenty of

K2V6].

29. *Donkeys, Rabbit and Dogs and Elephants (18th livestock census)*, OPEN GOV’T DATA (OGD) PLATFORM INDIA (Nov. 16, 2015), <https://data.gov.in/catalog/donkeys-rabbit-and-dogs-and-elephants-18th-livestock-census> [<https://perma.cc/76QP-EX6K>] (providing data on several animals, including totals by species and breed at both the state and district levels).

30. Cass R. Sunstein, *Output Transparency vs. Input Transparency* 9-13, <https://ssrn.com/abstract=2826009> [<https://perma.cc/F7Y7-AEGV>] (May 25, 2017).

31. When Data.gov launched in May 2009, it made forty-seven datasets searchable; that number skyrocketed to 130,000 by 2015. Jasamyn Roberts, *Data.gov Turns Six!*, DATA.GOV (May 21, 2015), <https://www.data.gov/meta/data-gov-turns-six> [<https://perma.cc/263J-46H8>]. As of this writing, Data.gov contains over 193,000 published datasets. DATA.GOV, *supra* note 22.

32. See, e.g., Harlan Yu & David G. Robinson, *The New Ambiguity of “Open Government,”* 59 UCLA L. REV. DISC. 178 (2012) (discussing ambiguities in the term “open government”).

unexpected riches.”³³

C. Differences in Audience

Open data also assumes a broader audience than FOIA. FOIA was written by and for journalists.³⁴ Yet, interestingly, corporations—knowing what to look for, knowing where to look, and having the resources to navigate the complex processes of filing requests—ultimately became the overwhelming and primary users of FOIA.³⁵ In a further departure from FOIA’s journalistic origins, many current FOIA requestors attempt to hobble the machinery of the administrative state through nuisance requests.³⁶

But open data anticipates—and thus far has attracted—a diverse and less consistently corporate audience. Unlike responses to FOIA requests, open data is directed to a wider public and is published for all to reuse. Beneficiaries of this public character include computer programmers and data scientists with the skills to draw insights from the data; academic users seeking information as the basis for original research, especially empirical social science about policymaking; and commercial users looking to create new products and services. This public direction and benefit of open data can be seen in practice. For example, the New York City Mayor’s Office used municipal open data to stimulate entrepreneurship. Through its *Business Atlas*, the City provided small enterprises with the business intelligence they need to know where to open new restaurants or shops.³⁷

Precisely because realizing the value of open data depends on collaboration with those willing to add value to it, the open data ecosystem is populated by wider audiences with different incentives from corporate users. In some cases, for-profit companies are working side by side with nonprofits to use the data as a core asset to create data-driven products and services. One example of this is BrightScope, which worked with previously “locked up” Department of Labor Form 5500 retirement plan data to offer better decision-making tools to investors.³⁸ As the founders describe it:

33. Rufus Pollock, *The Open Data Census – Tracking the State of Open Data Around the World*, OPEN KNOWLEDGE FOUND.: BLOG n.2 (Feb. 20, 2013), <http://blog.okfn.org/2013/02/20/open-data-census-tracking-the-state-of-open-data-around-the-world> [https://perma.cc/6C9P-ZWSL].

34. Patricia Wald, *The Freedom of Information Act: A Short Case Study in the Perils and Payback of Legislating Democratic Values*, 33 EMORY L. J. 649, 650-52 (1984).

35. Pozen, *supra* note 8, at 1103.

36. *Id.* at 1103-04.

37. CITY OF NEW YORK, *NYC Business Atlas* (2017), <https://maps.nyc.gov/businessatlas> [https://perma.cc/UD5A-S4RB].

38. Ryan Alfred & Mike Alfred, *From Entrepreneurs to Civic Entrepreneurs*, in BEYOND TRANSPARENCY 65, 72 (Brett Goldstein ed., 2013), <http://beyondtransparency.org/pdf/BeyondTransparency.pdf> [https://perma.cc/P745-NDT6].

While BrightScope started with DOL data, as we have grown we have gathered data and information from a variety of public sources, including the Securities and Exchange Commission (SEC), the Census Bureau, and the Financial Industry Regulatory Authority (FINRA). Through the process of identifying high-value datasets and integrating them into our databases, we have encountered all different types of public disclosure.³⁹

In contrast to FOIA, one of the biggest users of open data has been government itself, including officials wishing to make use of their own data to improve how they deliver services and make policy. For example, the Centers for Medicare and Medicaid Services (CMS) use their own billing and payment data to improve service delivery and reduce costs.⁴⁰ In addition, Chicago's city government used its data on restaurant inspections to create an algorithm to predict food-safety violations.⁴¹ This project increased the effectiveness of its inspections by 25%.⁴² By giving government actors access to more and better data, and especially by giving state and local government access to the same data as federal officials, the open data movement allows comparisons across jurisdictions and unlocks new, more innovative regulatory approaches. When the federal government ceases to have a monopoly on the data, it calls into question who is in the best and most informed position to regulate and opens opportunities for decentralized regulation.

II. THE OPEN DATA MOVEMENT: THE FUEL FOR DATA-DRIVEN ACTIVISM

The concept of open access to scientific research to enable replication of research findings and scholarly collaboration long predates open data. Simon Chignard traces this history of open scientific data for the common good back to until at least the early 1940s.⁴³ Robert Merton, an important figure in the sociology of science, believed that the results of research must be freely accessible to all.⁴⁴ However, the development of new "information technologies [gave] new breath to this philosophy of commons."⁴⁵

Contemporary policies of open data are only possible now because of the evolution of the technologies of "big data."⁴⁶ New collection, storage,

39. *Id.*

40. OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, ANALYTICAL PERSPECTIVES, BUDGET OF THE UNITED STATES GOVERNMENT, FISCAL YEAR 2016, 67, <https://obamawhitehouse.archives.gov/sites/default/files/omb/budget/fy2016/assets/spec.pdf> [<https://perma.cc/5DJ2-X8X2>].

41. Beth Simone Noveck, *Five Hacks for Digital Democracy*, 544 NATURE 287, 288 (2017).

42. *Id.*

43. Simon Chignard, *A Brief History of Open Data*, PARIS INNOVATION REV. (Mar. 29, 2013), <http://parisinnovationreview.com/2013/03/29/brief-history-open-data> [<https://perma.cc/6RU2-HBYD>].

44. *Id.*

45. *Id.*

46. Big data is the buzz word of this current era. See Solon Barocas & Andrew D. Selbst, *Big*

transmission, visualization, and analytic technologies have triggered a massive proliferation of detailed, often highly-individualized datasets about everything from health and wellness to phone and purchase records. Add to the volume of data mobile phones and sensors that collect the data that people gather through crowdsourcing activities (sometimes called citizen science), and the volume of big data is exploding.⁴⁷ The McKinsey Global Institute reports that the volume of data worldwide is growing exponentially,⁴⁸ and this trend shows no sign of slowing. In late 2012, the International Data Corporation, a market intelligence firm, estimated that the world contained some 2.8 zettabytes (2.8 trillion GB) of data,⁴⁹ a figure expected to grow more than ten-fold by 2020.⁵⁰

Of course, just because data is collected does not mean that it is used or that it even gets in the hands of those who wish to use it. It is open data laws and policies that provide the key for unlocking big data's potential and inviting collaboration. Open data transforms information collected by government pursuant to statutory or regulatory mandates into the raw material for problem-solving and the creation of useful tools in the public interest.

A. History of U.S. Federal Initiatives in Open Data

On his first day in office in 2009, fulfilling a campaign promise made in 2007, President Obama signed the Memorandum on Transparency and Open Government, declaring that “[i]nformation maintained by the Federal Government is a national asset,” and calling for the use of “new technologies to put information about [agency] operations and decisions

Data's Disparate Impact, 104 CAL. L. REV. 671, 673 (describing big data as the buzz word of the age); see also *Big Data: A Tool for Inclusion or Exclusion? Understanding the Issues*, FED. TRADE COMM'N 1 (Jan. 2016), <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> [<https://perma.cc/9JUD-QBEZ>].

47. *Data, Data Everywhere*, THE ECONOMIST (Feb 25, 2010), <http://www.economist.com/node/15557443> [<https://perma.cc/W4HL-459J>] (“The world contains an unimaginably vast amount of digital information which is getting ever vaster ever more rapidly.”)

48. See James Manyika et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, 16-18, MCKINSEY GLOBAL INST. (Oct. 2013), <http://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation> (click on “Full Report (PDF – 6MB)”) [<https://perma.cc/Z7SA-79KD>].

49. John Gantz & David Reinsel, *The Digital Universe in 2020: Big Data, Bigger Digital Shadows, and Biggest Growth in the Far East*, IDC iVIEW 15 (Dec. 2012), <https://www.emc.com/collateral/analyst-reports/idc-the-digital-universe-in-2020.pdf> [<https://perma.cc/3G4G-LSZS>].

50. John Murdoch-Burn, *Study: Less than 1% of the World's Data Is Analyzed, over 80% Is Unprotected*, THE GUARDIAN (Dec. 19, 2012), <http://www.theguardian.com/news/datablog/2012/dec/19/big-data-study-digital-universe-global-volume> [<https://perma.cc/BX9P-VRSD>]. For another method of understanding this explosion, see M.G. Siegler & Eric Schmidt, *Every 2 Days We Create as Much Information as We Did Up to 2003*, TECHCRUNCH (August 4, 2010), <http://techcrunch.com/2010/08/04/schmidt-data> [<https://perma.cc/9MGW-8678>].

online and [to make it] readily available to the public.”⁵¹

When the federal government’s open data repository, called Data.gov, launched in May 2009, it made forty-seven datasets searchable,⁵² turning the principles of the Memorandum into practice by creating a tangible and central place for agencies to list government data and for the public to find it.⁵³ Later that year, the OMB directed federal agencies to release not only data about the workings of government but also “high-value” information.⁵⁴ The choice to broaden the forty-year-old definition of government transparency responded to what both the technologies of big data and the technologies of collaboration could make possible today. In other words, the directive emphasized the broad public benefits and the need to disclose new kinds of government information as open data, such as locations of reported crimes, weather information, and information that could foster new businesses.

The Obama White House open data policy was part of a broader set of open government mandates⁵⁵ that called for agencies to inventory what information they collected⁵⁶ and to move—although with no definitive deadlines for completion—toward the proactive publication of classes of information in their entirety, such as air and water quality measures, safety records, or visitor logs.

In 2013, the federal government recommitted to its open data policy by issuing an Executive Order on “Making Open and Machine Readable the New Default for Government Information” to advance and accelerate open data implementation in federal agencies.⁵⁷ The Order reiterated the utilitarian and instrumentalist underpinnings of the earlier policies by stating explicitly that “[o]penness in government strengthens our democracy, promotes the delivery of efficient and effective services, and contributes to economic growth.”⁵⁸ The Order cites as examples the government’s release of both weather data⁵⁹ and geo-locational data,⁶⁰

51. Memorandum for the Heads of Exec. Dep’ts and Agencies on Transparency and Open Government, 2009 DAILY COMP. PRES. DOC. 1 (Jan. 21, 2009); see also *Experts Praise Barack Obama’s Technology and Innovation Agenda*, OBAMA FOR AM. (Nov. 14, 2007), <https://my.ofa.us/page/-/press/111407%20Tech%20innovation%20experts%20doc.pdf> [<https://perma.cc/V57S-D5FV>].

52. Roberts, *supra* note 31.

53. See *About*, DATA.GOV, <https://www.data.gov/about> [<https://perma.cc/9K83-ST6Z>] (“Data.gov is primarily a federal open government data site. . . . Data.gov does not host data directly, but rather aggregates metadata about open data resources in one centralized location.”).

54. See Orszag, *supra* note 23, at 7-8 (defining high-value datasets).

55. See *id.* at 3 (listing some presidential open government initiatives).

56. *Id.* at 7-8.

57. Exec. Order No. 13642, 3 C.F.R. § 13642 (2013).

58. *Id.*

59. *Id.* For further reading on this topic, see Christina Rogawski, Stefaan Verhulst & Andrew Young, *NOAA Open Data Portal: Creating a New Industry Through Access to Weather Data*, GOVLAB & OMIDYAR NETWORK (2016), <http://odimpart.org/files/case-studies-noaa.pdf> [<https://perma.cc/S4UD-Y8CT>].

60. 3 C.F.R. § 13642. For further reading on this topic, see Andrew Young, Christina Rogawski & Stefaan Verhulst, *United States GPS System: Creating a Global Public Utility*,

which enabled weather apps and GPS devices, respectively. Entrepreneurship and innovation—rather than accountability—are emphasized: “As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans’ lives and contributes significantly to job creation.”⁶¹

The need to populate Data.gov with high-value datasets has spurred agencies to automate the process of inventorying the data stored on their servers without needing to know who created or currently maintains the data. In turn, this automation has further accelerated the discovery of datasets to which Data.gov can then point.⁶²

Further laws have followed, broadening the scope of data covered under open data statutes and policies. The Digital Accountability and Transparency Act (DATA), signed into law in 2014, calls for publishing all federal government spending data as open data in standardized formats by 2017.⁶³ In late 2016, the Senate unanimously passed the “Open, Public, Electronic, and Necessary Government Data Act” or the “OPEN Government Data Act,”⁶⁴ which calls for inventorying and publishing all government information as open data.⁶⁵ The Congressional Budget Office scored the cost of the legislation as “negligible,”⁶⁶ and supporters re-introduced the bill in late March 2017.⁶⁷

The re-introduction of this law speaks to the persistent popularity of open data as a new tool of policymaking. In addition to this supply-side push, an increasing demand for data to support efficient, evidence-based practices in government has spurred the popularity of open data. The authors of *Moneyball for Government* describe the goals of this movement as follows:

- Building evidence about the practices, policies and programs that will achieve the most effective and efficient

GOVLAB & OMDIYAR NETWORK (2016), <http://odimpact.org/files/case-studies-gps.pdf> [<https://perma.cc/RY6R-ZUVH>].

61. 3 C.F.R. § 13642.

62. Roberts, *supra* note 31.

63. Digital Accountability and Transparency (DATA) Act of 2014, 31 U.S.C. § 6101.

64. 162 CONG. REC. S7135 (daily ed. Dec. 9, 2016).

65. OPEN Government Data Act, S. 2852 114th Cong. (2016). The provisions of the OPEN Government Data Act now constitute Title II of the Foundations for Evidence-Based Policymaking Act of 2017. H.R.4174, 115th Cong. (2017). The House passed this bill on November 15, 2017. 163 CONG. REC. H9291 (daily ed. Nov. 15, 2017).

66. S. 2852, *OPEN Government Data Act*, CONG. BUDGET OFFICE (Dec. 5, 2016), <https://www.cbo.gov/sites/default/files/114th-congress-2015-2016/costestimate/s2852.pdf> [<https://perma.cc/J3BC-T3VM>].

67. See 163 CONG. REC. H2557 (daily ed. Mar. 29, 2017) (introducing “H.R. 1770. A bill to expand the Government’s use and administration of data to facilitate transparency, effective governance, and innovation, and for other purposes; to the Committee on Oversight and Government Reform.”); 163 CONG. REC. S2099 (daily ed. Mar. 29, 2017) (introducing “S. 760. A bill to expand the Government’s use and administration of data to facilitate transparency, effective governance, and innovation, and for other purposes; to the Committee on Homeland Security and Governmental Affairs.”).

results so that policymakers can make better decisions;

- Investing limited taxpayer dollars in practices, policies and programs that use data, evidence and evaluation to demonstrate they work; and
- Directing funds away from practices, policies, and programs that consistently fail to achieve measurable outcomes.⁶⁸

In the United States, this agenda appeals to both the right and the left politically;⁶⁹ presumably the former sees open data as a pathway to smaller, more efficient government, while the latter uses open data as a tool to pursue more effective and evidence-based social programs. The bipartisan interest in evidence-based approaches to governing has fueled demand for more access to administrative information of all kinds, including the data that agencies collect about companies, workplaces, and the environment.⁷⁰

B. Other Governments, Corporations, Civil Society: The Open Data Movement Takes Hold

Open data's popularity has not solely been a U.S. phenomenon. Indeed, open data has begun to take hold around the world. Since 2011, seventy-five countries have signed onto the Open Government Partnership Declaration.⁷¹ These governments have agreed to "pro-actively provide high-value information, including raw data, in a timely manner, in formats that the public can easily locate, understand and use, and in formats that facilitate reuse."⁷² The Open Knowledge Foundation Global Open Data Index lists ninety-four countries with open data initiatives.⁷³ Fifteen

68. *Moneyball Principles*, MONEYBALL FOR GOV'T, <http://moneyballforgov.com/moneyball-principles> [<https://perma.cc/H9HA-GFC4>].

69. The OPEN Government Data Act was, in both the House and the Senate, jointly reintroduced by a Democrat and a Republican. See 163 CONG. REC. H2557 (daily ed. Mar. 29, 2017); 163 CONG. REC. S2099 (daily ed. Mar. 29, 2017).

70. See, e.g., Evidence-Based Policymaking Commission Act of 2016, Pub. L. No. 114-140, 130 Stat. 317 (creating a commission to study the use of open government data to conduct program evaluation). This bill was co-written and sponsored by Republican Congressman Paul Ryan and in Democratic Senator Patty Murray. See *About CEP*, COMM'N ON EVIDENCE-BASED POLICYMAKING (July 10, 2017), <https://www.cep.gov/about.html> [<https://perma.cc/3GMK-9L9R>] (noting that Ryan and Murray sponsored the bill); Press Office, *Evidence-Based Policy Commission Gets to Work*, SPEAKER PAUL RYAN (July 26, 2016), <http://www.speaker.gov/press-release/evidence-based-policy-commission-gets-to-work> [<https://perma.cc/8F9Z-DD74>] (noting that Ryan and Murray co-authored the bill).

71. *Open Government Declaration*, OPEN GOV'T P'SHIP, <http://www.opengovpartnership.org/open-government-declaration> [<https://perma.cc/7ZVS-YDGG>].

72. *Id.*

73. *Global Open Data Index*, OPEN KNOWLEDGE INT'L, <https://index.okfn.org/place> [<https://perma.cc/PZK7-XNF6>].

countries have adopted the International Open Data Charter, which is even more aggressive, calling for publication of all government data in digital formats by default.⁷⁴

In parallel, over 436 partners from national governments and non-governmental, international, and private sector organizations have agreed to a joint Statement of Purpose on using open data to solve long-standing agricultural problems, to the benefit of farmers and consumers alike.⁷⁵ This effort, known as the Global Open Data for Agriculture and Nutrition (GODAN), “supports the proactive sharing of open data to make information about agriculture and nutrition available, accessible and usable to deal with the urgent challenge of ensuring world food security.”⁷⁶

Although the quantity of available data is exploding, it is not necessarily the kind of data traditionally thought of as “FOIA data,” the release of which would, arguably, lead to greater government accountability and a decrease in the reality and perception of corruption. For example, while the United Kingdom, Canada, Greece, and Uruguay have leapt ahead of other countries in publishing government spending data (data which is, in fact, traditionally gathered through FOIA requests),⁷⁷ no other countries have yet followed their lead.⁷⁸

The spread of open data is not limited to government alone. In much the same way as government’s investment in the internet jumpstarted private sector investment in technology,⁷⁹ government innovation in the standards and policies of open data seems to be catalyzing follow-on investment by private industry in opening its own data for private gain and public good.⁸⁰ Companies have even begun to engage in “corporate data philanthropy,” whereby firms publish datasets, such as anonymized customer ride sharing data, for outside collaboration.⁸¹

Such corporate data philanthropy is the next wave in corporate social responsibility. Much of the data needed for solving social problems resides with the private sector—in the form of, for instance, web clicks, online

74. See *Adopted By*, OPEN DATA CHARTER, *supra* note 9.

75. *Partners*, GODAN, <http://www.godan.info/partners> [<https://perma.cc/VN8B-KAER>].

76. *About* GODAN, GODAN, <http://www.godan.info/pages/about-godan> [<https://perma.cc/YJV4-J4E6>].

77. *Open Data Barometer: ODB Global Report*, WORLD WIDE WEB FOUND. 24 (4th ed. 2017), <http://opendatabarometer.org/doc/4thEdition/ODB-4thEdition-GlobalReport.pdf> [<https://perma.cc/P2LT-FNVG>].

78. *Id.*

79. Mariana Mazzucato, *Opinion, State of Innovation: Busting the Private-Sector Myth*, THE NEW SCIENTIST (Aug. 21, 2013), <https://www.newscientist.com/article/mg21929310-200-state-of-innovation-busting-the-private-sector-myth> [<https://perma.cc/WS9P-C3LC>].

80. Beth Simone Noveck, *Data Collaboratives: Sharing Public Data in Private Hands for Social Good*, FORBES (Sept. 24, 2015), <https://www.forbes.com/sites/bethsimonenoveck/2015/09/24/private-data-sharing-for-public-good> [<https://perma.cc/5D6D-A3Q4>].

81. See, e.g., Douglas MacMillan, *Uber Offers Trip Data to Cities, Starting with Boston*, WALL ST. J.: DIGITS (Jan. 13, 2015), <https://blogs.wsj.com/digits/2015/01/13/uber-offers-trip-data-to-cities-starting-in-boston> [<https://perma.cc/ZBG8-M9UD>].

purchases, sensor data, and call data records.⁸² Many recent examples illustrate this philanthropic wave. Satellite companies regularly share data with public agencies to augment crisis response and give officials the real-time data needed to coordinate relief efforts.⁸³ Twitter donated data to the University of Wollongong in Australia to enable the creation of a publicly accessible real-time map of flood conditions in Jakarta, allowing Jakarta's residents to monitor floods and create individual emergency plans.⁸⁴ In the Ivory Coast and in Senegal, Orange Telecom anonymized customer call data and handed it to researchers who used the data to predict how waterborne parasites and disease travel.⁸⁵ Many other "data collaboratives" have also taken shape, involving participants from private companies, research institutions, government agencies, and other organizations, united in the need to exchange data to help solve public problems.⁸⁶

International organizations, too, have jumped on the bandwagon, with both the World Bank and the United Nations committing to making their own data available as open data.⁸⁷ The United Nations views open data as a key tool in the effort to achieve and measure progress toward the Sustainable Development Goals (SDGs), the human rights and development agenda adopted by 130 countries in 2016.⁸⁸ Between now and 2030, these countries are committing to seventeen calls to action, namely:

to end poverty and hunger everywhere; to combat inequalities within and among countries; to build peaceful, just and inclusive societies; to protect human rights and promote gender equality and the empowerment of women and girls; and to ensure the lasting protection of the planet and its natural resources. We resolve also to create conditions for sustainable, inclusive and sustained economic growth, shared prosperity and decent work for all, taking into

82. Stefaan Verhulst, *Corporate Social Responsibility for a Data Age*, STAN. SOC. INNOVATION REV. (Feb. 15, 2017), https://ssir.org/articles/entry/corporate_social_responsibility_for_a_data_age [<https://perma.cc/8GAQ-2CG3>].

83. Matt Stempeck, *Sharing Data Is a Form of Corporate Philanthropy*, HARV. BUS. REV. (July 24, 2014), <https://hbr.org/2014/07/sharing-data-is-a-form-of-corporate-philanthropy> [<https://perma.cc/5QY3-JJ8M>].

84. Tomas Holderness & Etienne Turpin, *How Tweeting About Floods Became a Civic Duty in Jakarta*, THE GUARDIAN (Jan. 26, 2016), <https://www.theguardian.com/public-leaders-network/2016/jan/25/floods-jakarta-indonesia-twitter-petajakarta-org> [<https://perma.cc/WH59-JNP7>].

85. Vincent D. Blondel et al., *Data for Development: The D4D Challenge on Mobile Phone Data*, ARXIV (Sept. 29, 2012), <https://arxiv.org/pdf/1210.0137.pdf> [<https://perma.cc/RL3D-R7BV>].

86. See *Data Collaboratives*, GOVLAB, <http://www.datacollaboratives.org> [<https://perma.cc/EVC8-D68J>].

87. See, e.g., U.N. OFFICE OF THE HIGH COMMISSIONER FOR HUMAN RIGHTS, GOOD GOVERNANCE PRACTICES FOR THE PROTECTION OF HUMAN RIGHTS, at 37, U.N. Doc. HR/PUB/07/4, Sales No. E.07.XIV.10 (2007) (discussing the role of a visual and easily consumable budget dataset to ensure more equitable access to social services in Ecuador).

88. See G.A. Res. 70/1, ¶ 3 (Sept. 25, 2015).

account different levels of national development and capacities.⁸⁹

The Open Data Barometer aptly summarizes the relationship between open data and the SDG targets: “Open data is essential to building accountable and effective institutions, and to ensuring public access to information—both goals of SDG 16.”⁹⁰ As of April 2017, reports UNDP, 84 of 230 SDGs indicators have no methodologies for measuring progress.⁹¹ Understanding the status of the 2030 Agenda for Sustainable Development and progress toward its laudable goals requires access to official sources of data to feed analytical models. UNDP’s Innovation Lab and others are campaigning for use of unofficial sources of data as well such as information collected by citizens rather than governments where public institutions, especially in the developing world, do not collect adequate administrative data.⁹² But the first step is to open government data where it does exist to create accountability. In this context, it comes as no surprise that the *Open Data Barometer* concluded that “we need robust data to drive democracy and development—and a lot of it.”⁹³

The World Bank has also acknowledged the usefulness of open data in achieving the SDGs:

Open Data can help countries around the world address a wide range of development goals. It is an essential, versatile resource that can increase trust in government, boost economic growth, create jobs, and improve essential products and services. As the Sustainable Development Goals are being set, countries across the globe can look to Open Data as a ready means to help meet them. Open Data can help improve food security, healthcare, education, cities, the environment, and other public and private services that are critical to development.⁹⁴

In sum, there exists significant support for open data in the SDG context by major international organizations and by traditional data advocates alike. The Global Partnership for Sustainable Development Data,⁹⁵ for instance, campaigns for better use of data, generally, and has

89. *Id.*

90. *Open Data Barometer: ODB Global Report*, *supra* note 77, at 6.

91. Jennifer Colville, Presentation to MacArthur Foundation Research Network: Sustainable Development Goals Tier III Indicators (June 14, 2017) (presentation on file with author).

92. *Id.*

93. *Open Data Barometer: ODB Global Report*, *supra* note 77, at 6.

94. World Bank Group, Policy Note: Open Data for Sustainable Development, at 23, W.B. Doc. ICT01 (2015), <http://pubdocs.worldbank.org/en/741081441230716917/Open-Data-for-Sustainable-development-PN-FINAL-ONLINE-September1.pdf> [<https://perma.cc/6WZU-ANDH>].

95. See *SDG Data Revolution Roadmaps Toolbox: Open Data Module*, GLOB. P'SHIP FOR SUSTAINABLE DEV. DATA 1 (July 16, 2016), <https://static1.squarespace.com/static/55f7418ce4b0c5233375af19/t/57911d77abebafb816e24f4>

taken an interest in open data's usefulness for accomplishing and evaluating progress toward the SDGs.⁹⁶ Such organizational interest is not limited to the SDGs, however; a growing alphabet soup of NGOs advocate for open data either as an end unto itself or as a tool to accomplish other goals. Among these open data campaigners and supporters are the Open Data Institute, the Open Knowledge Foundation, Open Data Watch, Data Pop Alliance, and the World Wide Web Foundation.⁹⁷

III. FROM DATA TO ACTION: OPEN DATA'S IMPACTS

A. Analytical Opportunities

Identifying Inequities and Inefficiencies by Measuring Past Performance

The technologies of big data enable more evidence-based practices. By analyzing data that government generates, collects, and shares, policymakers can understand *past* performance of public policies and services, evaluating both their efficiency and how they impact different populations. For example, the Society for Cardiothoracic Surgery (SCTS) and the National Health Service (NHS) in the United Kingdom have published a decade's worth of data on heart surgery outcomes, together with an app that allows the public to analyze and make sense of the results. Research shows that open publication of the data has led to a thousand fewer deaths and cost savings system-wide.⁹⁸

Targeting Allocation of Scarce Resources

Larger quantities of data enable the delivery of more tailored interventions in the *present* by making it possible, for example, to identify and divert someone suffering from mental illness or substance abuse away from the criminal justice system and towards appropriate treatment.⁹⁹ By using their own data better, governments at every level can match people to the benefits to which they are entitled. For example, Louisiana's Department of Health uses Supplemental Nutrition Assistance Program

58/1469128061047/DRAFTOpenDataSDGToolboxModuleJuly2016.pdf
[<https://perma.cc/25TT-QAS7>].

96. Joel Gurin, Laura Manley & Audrey Ariss, *Sustainable Development Goals and Open Data*, WORLD BANK GROUP: INFO. AND COMMUNICATIONS FOR DEV. BLOG (Sept. 25, 2015), <http://blogs.worldbank.org/ic4d/sustainable-development-goals-and-open-data> [<https://perma.cc/Y6AS-VVXY>].

97. *Commitments*, GLOB. P'SHIP FOR SUSTAINABLE DEV. DATA, <http://www.data4sdgs.org/commitments/> [<https://perma.cc/B32K-P7VT>].

98. Stefaan Verhulst et al., *The Open Data Era in Health and Social Care*, GOVLAB 30 (2014), <http://www.thegovlab.org/static/files/publications/nhs-full-report.pdf> [<https://perma.cc/5RKK-BUQ3>].

99. One such project tailored to criminal justice is the Data Justice Network. See DATA JUSTICE NETWORK, <https://www.datajustice.us> [<https://perma.cc/D3HS-UPJG>].

(SNAP) enrollment data to sign people up for health benefits.¹⁰⁰ Out of nearly 900,000 SNAP recipients, Louisiana has proactively enrolled 105,000 in Medicaid without a separate application process, instead using a four question yes-no survey.¹⁰¹ This approach has helped some of the state's poorest residents to get the benefits to which they were entitled, while also saving the state approximately \$1.5 million in administrative costs.¹⁰²

Predicting Outcomes

Better access to data even helps with foretelling *future* outcomes, such as predicting who is likely to be a frequent visitor to the emergency room, thereby enabling more targeted interventions and treatment.¹⁰³ Having access to data and predictive models could help answer questions like: how many people will a disease potentially infect; how far and how quickly will the disease spread; what areas and people are at highest risk; and when are they most at risk. It would also help government understand how to best make use of limited resources to prevent and deal with outbreaks and stem the spread of such pandemics as Zika or dengue.¹⁰⁴

B. The Social Impacts of Open Data

Like the analytical uses of open data, the potential social impacts of opening data are myriad and multiplying. These benefits fall into five categories. Although the first such category relates to the traditional FOIA role of holding government to account, the other four extend to a wide range of additional ways that open data can aid consumers and communities.

Improving Governmental Accountability

Open data often better achieves the goal for which FOIA was created, namely greater government accountability. For example, in Brazil, open

100. *Louisiana Receives Approval for Unique Strategy To Enroll SNAP Beneficiaries in Expanded Medicaid Coverage*, LOUISIANA DEP'T. OF HEALTH (June 1, 2016), <http://dhh.louisiana.gov/index.cfm/newsroom/detail/3838> [https://perma.cc/SAK8-X4SK].

101. *Id.*; see also Dep't of Health and Hospitals, *Sample of SNAP Offer Letter*, LOUISIANA DEP'T OF HEALTH (May 2, 2016) (showing an example of the SNAP offer letter and survey questions sent to targeted populations in Louisiana).

102. *Louisiana Receives Approval for Unique Strategy to Enroll SNAP Beneficiaries in Expanded Medicaid Coverage*, *supra* note 100.

103. Noreen Kamal et al., *Addressing Emergency Department Overcrowding Through a Systems Approach Using Big Data Research*, 5 J. HEALTH & MED. INFORMATICS 3 (Feb. 18, 2014); see also Stefaan Verhulst et al., *supra* note 98, at 28.

104. Danielle Barron, *Is Artificial Intelligence Key to Dengue Prevention?*, BREAK DENGUE BLOG (June 8, 2016), <https://www.breakdengue.org/artificial-intelligence-dengue-prevention> [https://perma.cc/MXG5-83QJ].

data facilitated the creation of the Portal de Transparencia,¹⁰⁵ a set of online tools for exploring the federal budget, that attracts 900,000 discrete visitors every month¹⁰⁶ and has led to a 25% decrease in credit card spending by public servants.¹⁰⁷ At the local level, open data drives various open “checkbook” websites. The City of New York¹⁰⁸ and hundreds of municipalities in Austria¹⁰⁹ have made their spending data more transparent and easy to visualize online. Similarly taking advantage of the whole-dataset approach of open data, the civil society group Mazdoor Kisan Shakti Sangathan (MKSS) paints government budgets on the walls of villages in Rajasthan, India and holds theatrical readings of the information to enable the communities to point out “bridges built to nowhere.”¹¹⁰

Such disclosure is not guaranteed to have a substantive effect. As with most reforms, the right conditions must be in place first. Specifically, where such disclosure is a positive act undertaken as a collaboration between government and civil society or where the public otherwise has both the interest and the tools to hold government to account, the traditional gap between disclosure and accountability narrows. The growing subfield of open contracting—led by the international Open Contracting Partnership—is further strengthening the connection between disclosure and accountability by making public procurement information available as open data.¹¹¹ The Ukraine’s open contracting platform, Prozorro.org, has enabled outside organizations to investigate fraudulent contracts. These reforms have saved Ukraine’s taxpayers more than U.S. \$300 million.¹¹²

Improving Accountability of Private Actors

Although shining a light on government is critically important, the potential impact of open data is broader than FOIA’s goals of improving government transparency.¹¹³ Among its many applications, open data can

105. PORTAL DA TRANSPARÊNCIA DE BRASIL [BRAZIL TRANSPARENCY PORTAL], <http://www.portaltransparencia.gov.br> [<https://perma.cc/MXG5-83QJ>].

106. Auralice Graft, Stefaan Verhulst & Andrew Young, *Brazil’s Open Budget Transparency Portal: Making Public How Public Money Is Spent*, GOVLAB & OMIDYAR NETWORK 10 (2016), <http://odimpact.org/files/case-study-brazil.pdf> [<https://perma.cc/38S9-WQG8>].

107. *Id.* at 11.

108. *Checkbook NYC*, NYC COMPTROLLER’S OFFICE, http://www.checkbooknyc.com/spending_landing/yeartype/B/year/118 [<https://perma.cc/F8JG-G2YP>].

109. See Bernhard Krabina, *Austria’s Award-winning Open Spending Portal Turns Two*, FISCAL TRANSPARENCY (Oct. 1, 2015), <http://www.fiscaltransparency.net/use/austrias-award-winning-open-spending-portal-turns-two> [<https://perma.cc/N3ZG-8HNR>].

110. See BETH SIMONE NOVECK, SMART CITIZENS, SMARTER STATE 10 (2015).

111. See *Global Principles*, OPEN CONTRACTING P’SHIP, <http://www.opencontracting.org/get-started/global-principles> [<https://perma.cc/X9UN-LYVW>].

112. Gavin Heyman, *Firing Up for 2017: What To Expect from Open Contracting*, OPEN CONTRACTING P’SHIP (Jan. 11, 2017), <http://www.open-contracting.org/2017/01/11/firing-up-2017> [<https://perma.cc/Z7JN-M27W>].

113. For more about the impacts of open data, see Stefaan Verhulst, *Open Data is Changing the World in Four Ways*, GOVLAB (Jan. 25, 2016), <http://thegovlab.org/open-data-is-changing-the-world-in-four-ways> [<https://perma.cc/44R6-CY89>].

also help hold private organizations and institutions to account. For example, when open data about organizations and about the products they produce is available in computable formats, governments and public watchdogs can “see” when a company is engaged in price gouging or can visualize a company’s labor and environmental violations on a map. In this way, regulators and the public “can get early warning signs about where oversight is most needed.”¹¹⁴

Open data can also help private actors self-regulate. For example, several states have begun collecting information about doctors’ patterns of prescribing opioid pain medication.¹¹⁵ By transparently showing doctors their own practices in comparison to those of other doctors (and, crucially, by doing so in a public way), open data catalyzes changes the behavior of less responsible prescribers who are wary of the public scrutiny that such practices might engender. This idea was recently piloted in Arizona, with participating counties experiencing a 10% reduction in opiate prescriptions and a 4% reduction in overdose deaths, as compared to a continued rise in opiate prescriptions in non-participating counties.¹¹⁶

Enhancing Consumer Choice Through Smart Disclosure

Open data can also empower consumers to make more informed decisions.¹¹⁷ By providing consumers more complete data about the cost, quality, and safety of the products and services they buy, as well as information about labor practices and health and environmental impacts,¹¹⁸ consumers can make more informed choices.

Smart disclosure has its roots in behavioral economics. It embodies two principles: (1) that better information will lead to better consumer decisions; and (2) that informing consumers will, in turn, make consumer markets more efficient, fostering competition and innovation. For example, although government-mandated labels have given new car buyers information on estimated miles per gallon for years, such labels were recently redesigned to make this information even more transparent and useful to consumers.¹¹⁹

With regard to open data, “consumer” should be understood broadly,

114. Joel Gurin & Beth Simone Noveck, *Corporations and Transparency: Improving Consumer Markets and Increasing Public Accountability*, in *TRANSPARENCY IN POLITICS AND THE MEDIA: ACCOUNTABILITY AND OPEN GOVERNMENT* 179, 181 (Nigel Bowles et al. eds., 2013).

115. Scott Calvert, *Doctors’ Individual Opioid Prescription ‘Report Cards’ Show Impact*, WALL ST. J. (Sept. 2, 2016), <http://www.wsj.com/articles/doctors-individual-opiate-prescription-report-cards-show-impact-1472856624> [<https://perma.cc/JA5L-XVT6>].

116. *Id.*

117. Note, however, that many of these benefits are still assumed. Because smart disclosure depends on the most recent technology to work, the approach is still nascent and requires further study.

118. For a good overview of smart disclosure, see Richard H. Thaler & Will Tucker, *Smarter Information, Smarter Consumers*, HARV. BUS. REV. (Jan-Feb. 2013), at 44-54.

119. Office of the Press Sec’y, *Obama Administration Finalizes Historic 54.5 MPG Fuel Efficiency Standards*, THE WHITE HOUSE (Aug. 28, 2012), <http://obamawhitehouse.archives.gov/the-press-office/2012/08/28/obama-administration-finalizes-historic-545-mpg-fuel-efficiency-standard> [<https://perma.cc/D74L-PTT9>].

as open data policies often encourage smart consumption that extends beyond traditional retail purchases. For example, in the United States, universities must disclose significant amounts of data in order to be eligible for federal higher education and financial aid programs.¹²⁰ The Department of Education then presents this data as part of its College Scorecard, a well-known and easy-to-read data display that helps parents and students make more informed decisions about their college educations.¹²¹

Often, private organizations overlay user-friendly tools on top of government-published open data. In Uruguay, for example, a nonprofit called DATA Uruguay used information from the Ministry of Health to create a website called A Tu Servicio.¹²² This website offers open access to “easily digestible, searchable and visualized infographics based on open government health data.”¹²³ The platform allows consumers to compare health care services and to make an informed choice “based on a wide range of parameters and indicators, such as facility type, medical specialty, care goals, wait times, and patient rights.”¹²⁴

The hope is that smart disclosure can result in more effective regulation that may, at the same time, be less burdensome to business than mandated and directive approaches such as rate regulation. But it is not yet well-understood whether and when a regulatory mandate to disclose information can substitute for more stringent price controls or other regulations.

Promoting Entrepreneurship

Open data’s impact often comes in the form of job creation and economic growth. For instance, publication of government contracts can boost public integrity and is also known to catalyze competition and innovation.¹²⁵ A study by the McKinsey Global Institute estimates that open data will contribute to the creation of \$3 trillion a year in global economic

120. See Andrea Sykes, *Information Required To Be Disclosed Under the Higher Education Act of 1965: Suggestions for Dissemination – A Supplemental Report*, NAT’L POSTSECONDARY EDUC. COOP. 1 (2011), <https://nces.ed.gov/pubs2012/2012831.pdf> [<https://perma.cc/8U7S-DRWF>] (noting that federal law mandates postsecondary institutional compliance with regard to “about 40 consumer information disclosures”).

121. *College Scorecard*, U.S. DEP’T OF EDUC., <https://collegescorecard.ed.gov> [<https://perma.cc/L9HE-ZT6E>]; see also *College Affordability and Transparency Center*, U.S. DEP’T OF EDUC., <http://collegecost.ed.gov/catc/Default.aspx> [<https://perma.cc/BBQ6-9E5X>]; *College Navigator*, U.S. DEP’T OF EDUC., <http://nces.ed.gov/collegenavigator> [<https://perma.cc/95XT-ZPW5>]; *Net Price Calculator Information Center*, U.S. DEP’T OF EDUC., NAT’L CENT. EDUC. STAT., <https://nces.ed.gov/ipeds/netpricecalculator> [<https://perma.cc/J9A6-YCY6>].

122. David Sangokoya et al., *URUGUAY’S A TU SERVICIO: EMPOWERING CITIZENS TO MAKE DATA-DRIVEN DECISIONS ON HEALTH CARE*, GOVLAB & OMIYAR NETWORK 3 (2016), <http://odimparct.org/files/case-study-uruguay.pdf> [<https://perma.cc/6MMS-B5V3>].

123. *Id.*

124. *Id.*

125. See *Why Open Contracting*, OPEN CONTRACTING P’SHIP, <http://www.open-contracting.org/why-open-contracting> [<https://perma.cc/93RD-LCQR>].

value.¹²⁶

The Open Data 500 is an original survey of thousands of companies worldwide that use open government data as a core business asset.¹²⁷ One of these companies is the New York-based startup Aidin, which uses government data about nursing home and home health provider comparisons to show patients the best available post-acute care providers.¹²⁸ Despite the fact that the data it uses to run its business is available to everyone, Aidin is able to do well by doing good. This is true even though the information that Aidin provides is, by definition, open to everyone. Although such success may appear contrary to conventional wisdom about business, research by Columbia professor Sheena Iyengar and postdoctoral research scholar Patrick Bergmann indicates that the absence of barriers to entry has largely not proven an impediment to investment in for-profit open data ventures.¹²⁹ Their analysis of the Open Data 500 network indicates that investors are just as likely to invest in companies based on non-rivalrous assets like open data as they are in companies using more traditional, proprietary data.¹³⁰

Saving Lives and Solving Problems

With available real-time data about emergencies, open data can help to make the delivery of government services far more effective. In the case of PulsePoint, open data saves lives. PulsePoint is an app that monitors a real-time feed of 911 calls and notifies qualified volunteers when there is a cardiac emergency nearby.¹³¹ This public-private collaboration, started by the former fire chief of San Ramon California, simply would not be possible without open data.¹³² Already, the app has been used in thousands of cardiac emergencies, saving countless lives.¹³³

126. James Manyika et al., *Open Data: Unlocking Innovation and Performance with Liquid Information*, MCKINSEY GLOB. INST. 6 (2013) (click on “Full Report (PDF - 7MB)”) [<https://perma.cc/597W-RBP5>].

127. *The OD500 Global Network*, GOVLAB: OPEN DATA 500 GLOBAL NETWORK, <http://www.opendata500.com> [<https://perma.cc/TX8L-DFYL>].

128. For a brief profile of Aidin, see *Aidin*, GOVLAB: OPEN DATA 500 GLOBAL NETWORK, <http://www.opendata500.com/us/aidin> [<https://perma.cc/B8R4-8CFG>].

129. Patrick Bergemann & Sheena Iyengar, *The Economic Impact of Open Data 4-9* (accepted into the October 2016 Open Data Research Symposium), <https://drive.google.com/file/d/0B4TpC6ecmrM7LUt5N1hJQTNQWm8/view> [<https://perma.cc/AK6Y-WCLP>].

130. *Id.*

131. Matt Wilson, *Pulsepoint App Turns Bystanders into First Responders*, MERCURY NEWS (June 11, 2014), <http://www.mercurynews.com/2014/06/11/pulsepoint-app-turns-bystanders-into-first-responders> [<https://perma.cc/99WZ-N2TW>].

132. See NOVECK, *supra* note 110, at 39-42 (describing Pulsepoint and the benefits of crowdsourcing, citizen engagement, and increasingly opening government).

133. See Beth Simone Noveck, *Swipe Right To Fix the World: Can Tinder-Like Tech Match Solutions to Problems?*, THE GUARDIAN (Dec. 29, 2015), <https://www.theguardian.com/commentisfree/2015/dec/29/swipe-right-to-fix-the-world-can-tinder-like-tech-match-solutions-to-problems> [<https://perma.cc/SYP5-23C3>].

IV. THE MORAL DATA ECONOMY: OPEN DATA, HUMAN RIGHTS, AND DEVELOPMENT

Whereas FOIA skews in practice toward businesses requesting competitor data,¹³⁴ open data is proving to be a vital collaborative tool for governments and civil society seeking to advance social justice. Such social change occurs in several ways: (a) by facilitating empirical social science, investigative journalism, and consumer protection, especially in domains in which data and quantification can help to reveal corruption and gross malfeasance;¹³⁵ (b) by advancing civil rights litigation by shining a light on disparate impacts that are often hiding in plain sight; (c) by assessing the effectiveness and fairness of our institutions of justice; (d) by reducing corruption in and enhancing the effectiveness of foreign development assistance and (e) by improving the monitoring and identification of human rights abuses.¹³⁶

A. Facilitating Empirical Social Science, Investigative Journalism, and Consumer Protection

Unlike a FOIA request, which typically yields extensive documents that must be painstakingly reviewed and analyzed,¹³⁷ open datasets are usually machine-readable and can be used to conduct analysis, draw comparisons, and spot individual problems quickly and with minimal human effort.

The promise of open data to accelerate such investigations is especially clear in the consumer protection realm. As consumer product offerings get more complex and confusing, regulatory mandates to open data can enable greater scrutiny and make analyzing patterns across an industry easier. For example, information disclosures make it possible to analyze those health insurance companies that deny the most claims or credit card companies making the most money off of late fees. In addition, applications based on open data could expose the practices of industries that rely on hidden fees, confusing terms of service, or other forms of obfuscation to profit from their customers' confusion.

134. Pozen, *supra* note 8, at 1103.

135. See Stefaan Baak, *Datafication and Empowerment: How the Open Data Movement Re-Articulates Notions of Democracy, Participation, and Journalism*, BIG DATA & SOC'Y 1 (2015).

136. See generally Noel Isama & Lindsay Ferris, *Using Open Data To Stimulate New Methods of Human Rights Monitoring*, SUNLIGHT FOUND.: BLOG (Feb. 25, 2016), <https://sunlightfoundation.com/2016/02/25/using-open-data-to-stimulate-new-methods-of-human-rights-monitoring> [<https://perma.cc/NWY4-DSC8>].

137. While FOIA requests are increasingly coming in digital formats, they still often require cataloging and individual searching in a way less conducive to rigorous review than open datasets. See Rebecca Williams, *Freedom of Information Tools, It's Time for an Open Data Inspired Upgrade*, SUNLIGHT FOUND.: BLOG (Aug. 24, 2014), <https://sunlightfoundation.com/2014/08/25/freedom-of-information-tools-its-time-for-an-open-data-inspired-upgrade> [<https://perma.cc/ZR4F-TJ5X>] (explaining problems in formatting and completeness of FOIA in Chicago, D.C., and New York, even when cities attempt to make part of the FOIA publicly available and in some sort of searchable format).

B. Advancing Civil Rights and Other Public Interest Litigation by Revealing Disparate Impacts

Increasingly, open data is an asset in civil rights litigation. Opening local government data about public works in Zanesville, Ohio revealed a fifty-year pattern of discriminatory water service provision.¹³⁸ Although the City of Zanesville's water line allowed the rest of Muskingum County access to clean water, the water line did not provide water to predominantly African-American neighborhoods of Zanesville, Ohio. Residents in these areas instead either had to use contaminated rainwater or transport water from water towers back to their homes.¹³⁹ The use of open data aided lawyers¹⁴⁰ in a successful civil rights lawsuit and jury award against Zanesville in 2008.¹⁴¹

To emphasize just how important this type of open data can be for facilitating cases and good outcomes without an overburdening of investigative resources, consider the difficulties that frustrated a recent case concerning suspicious charity activity. In May 2015, all fifty states and the District of Columbia, along with the Federal Trade Commission, filed a lawsuit against the Cancer Fund group of nonprofits and the individuals who run them.¹⁴² The suit alleged that the Cancer Fund group misappropriated \$187 million raised from American donors.¹⁴³ They confirmed their suspicion of fraud by scrutinizing federal nonprofit tax returns and state-by-state registration forms by hand.¹⁴⁴ If regulators had had fully open, machine-readable IRS Form 990 data, however, they would have likely found suspicious patterns earlier. Furthermore, with access to the entire corpus of such data, many more suits might have been brought and more consumers helped.¹⁴⁵

One could also imagine how much more effectively and efficiently

138. Christina Rogawski, Stefaan Verhulst & Andrew Young, *Ohio, USA: Kennedy v. City of Zanesville: Open Data as Evidence*, GOVLAB & OMIDYAR NETWORK 3 (2016), <http://odimpact.org/files/case-studies-us-kennedy.pdf> [<https://perma.cc/K7B7-4YQR>].

139. *Id.* at 3.

140. *Id.* at 6-8.

141. *Kennedy v. City of Zanesville*, 505 F.Supp. 2d 456 (S.D. Ohio 2007) (granting plaintiffs partial summary judgment). In July 2008, the jury awarded plaintiffs nearly \$11 million in damages. See Dirk Johnson, *For a Recently Plumbed Neighborhood, Validation in a Verdict*, N.Y. TIMES (Aug. 11, 2008), <http://www.nytimes.com/2008/08/12/us/12ohio.html> [<http://perma.cc/M55W-UR4A>].

142. See *FTC v. Cancer Fund of America, Inc.*, No. 2:15-CV-00884-NVW (D. Ariz. filed May 18, 2015); see also Beth Simone Noveck, *The Data That's Hiding in Plain Sight*, GOVERNING MAG. (June 3, 2015), <http://www.governing.com/columns/smart-mgmt/col-open-data-state-regulation-nonprofits-charities.html> [<https://perma.cc/XZ77-4A6X>] (discussing this lawsuit and how machine readable 990 forms likely would have made identification of suspicious activity easier).

143. See *Complaint, FTC v. Cancer Fund of America, Inc.*, No. 2:15-CV-00884-NVW, at 38-41 (D. Ariz. May 18, 2015), <https://www.ftc.gov/system/files/documents/cases/150519cancerfundcmpt.pdf> [<https://perma.cc/A7LV-MWW6>].

144. See Noveck, *supra* note 142.

145. *Id.*

Voting Rights Act litigation could proceed (or be avoided) if the Department of Justice were to compile data from each jurisdiction on its election standards, practices, and procedures in a machine-readable format. The Supreme Court's recent decision in *Cooper v. Harris*,¹⁴⁶ which affirmed that the North Carolina legislature impermissibly used race as a predominant factor in gerrymandering two districts, came almost six years after those district lines were drawn. With open data, it should be possible to automatically spot abridgements of the right to vote, thus helping to more quickly resolve voting rights cases.

C. Assessing the Effectiveness and Fairness of Institutions of Justice

Open data provides many opportunities to consider empirical reforms that address justice-related practices and institutions. One approach used by former President Barack Obama's administration was to convene more than fifty jurisdictions as part of a White House Police Data Initiative and to work with those jurisdictions to publish their anonymized policing data.¹⁴⁷ The even broader White House Data-Driven Justice Initiative (DDJI) partnered with 129 communities to bring about empirically-based reforms.¹⁴⁸ The DDJI focused both on making criminal justice data available as open data and encouraging secure sharing by government institutions, such as the criminal justice and mental health systems, of sensitive personal information. The emphasis was on using data to reduce unnecessary jail sentences among those charged with low-level, nonviolent misdemeanors, especially those in need of treatment for mental illness, substance use disorders, or chronic health problems.¹⁴⁹ The decentralized nature of the criminal justice system, with 18,000 law enforcement agencies,¹⁵⁰ 1,820 federal and state correctional facilities,¹⁵¹ and more than 3,000 local jails,¹⁵² means that the federal government alone does not have access to complete, or even superior, justice system data. Thus, the DDJI emphasized the need

146. 581 U.S. __ (2017).

147. Office of the Press Sec'y, *Fact Sheet: White House Police Data Initiative Highlights New Commitments*, THE WHITE HOUSE (Apr. 21, 2016), <https://obamawhitehouse.archives.gov/the-press-office/2016/04/22/fact-sheet-white-house-police-data-initiative-highlights-new-commitments> [<https://perma.cc/8FQA-7DQQ>].

148. Ron Davis et al., *Growing Number of Communities Are Using Data To Improve Policing and Criminal Justice*, THE WHITE HOUSE: BLOG (Oct. 13, 2016), <https://obamawhitehouse.archives.gov/blog/2016/10/13/growing-number-communities-are-using-data-improve-policing-and-criminal-justice> [<https://perma.cc/SAN6-DUR8>]. Unfortunately, the numbers are a far cry from the three thousand counties there are in the United States or the eighteen thousand police departments in the United States.

149. Barack Obama, *The President's Role in Advancing Criminal Justice Reform*, 130 HARV. L. REV. 811, 848 (2017).

150. Bureau of Justice Stat., *National Sources of Law Enforcement Employment Data*, U.S. DEP'T OF JUSTICE 1, NCJ 249681 (Apr. 2016), <https://www.bjs.gov/content/pub/pdf/nslead.pdf> [<https://perma.cc/SY2A-8FJL>].

151. DAVID W. NEUBAUER & HENRY F. FRADELLA, *AMERICA'S COURTS AND THE CRIMINAL JUSTICE SYSTEM* 7 (2013).

152. Office of the Press Sec'y, *supra* note 147.

for cross-jurisdictional data sharing to facilitate decentralized, distributed, and localized reform and innovation efforts. Miami-Dade County, Florida, for example, analyzed a combination of open data and non-open data (confidential, personally-identifiable data) from the county's health care and criminal justice systems and discovered that just ninety-seven people with serious mental illness accounted for \$13.7 million in services between 2010 and 2014.¹⁵³ Thanks to this empirical analysis, the county created targeted interventions that dropped costs by \$12 million.¹⁵⁴

Of course, open data is not limited to executive agency data. Courts are also opening their own data. The Court Statistics Project, for example, gathers and publishes caseload data from state courts of the fifty states, making it possible to see which states have the busiest dockets.¹⁵⁵ The Court Statistics Project is just one of an increasing number of projects that are using open data to evaluate how well legal institutions deliver services to the public. Argentina, Lithuania, and Slovakia also have judicial open data portals, which publish such datasets as judicial position openings, the CVs of applicants for those positions, average caseloads, and court budgets.¹⁵⁶

When such data is not readily available, civil society often takes the lead. The new Center on Civil Justice at NYU Law School is gathering and creating open data on class action lawsuits. It has paid special attention to awards in these cases, including who got what; what percentage went to the lawyers versus administration versus claimants; how it was distributed; how big the class was; and what kind of case it was.¹⁵⁷ In the long run, using aggregate data to analyze the performance of the civil justice system could revolutionize the way courts, juries, parties, lawyers, and others calculate damages. As Center Director Peter Zimroth writes:

Good data is essential for making good policy. Consumer class actions either compensate consumers or they do not. They either deter unwanted business practices or they do not. At the moment, the debate over these questions is being conducted with anecdote and personal experience. This is not helpful because the personal experiences of equally knowledgeable professionals often lead to opposite conclusions. Data exists to answer these questions. It is either in the possession of parties with little or no incentive to share it, or, when information is available in court files for some cases, it is prohibitively expensive to find the court file for those individual cases and then combine whatever information is available. The

153. *Id.*

154. *Id.*

155. *Court Statistics Project*, NAT'L CTR. FOR STATE COURTS, <http://www.courtstatistics.org> [<https://perma.cc/E9LX-4K7F>].

156. Slovakia's judicial open data portal is located at <https://otvorenesudy.sk> [<https://perma.cc/3H7P-NEJ5>]; Lithuania's judicial open data portal is located at <http://atvirasteismas.lt> [<https://perma.cc/36G2-GA8M>]; and Argentina's judicial open data portal is located at <http://datos.jus.gob.ar> [<https://perma.cc/P6HC-MKFT>].

157. *About*, CTR. ON CIVIL JUSTICE AT N.Y.U. SCH. OF LAW, <http://www.law.nyu.edu/centers/civiljustice/about> [<https://perma.cc/2FS2-Z69D>].

Center is beginning to explore ways that it can serve as a neutral repository for information that certainly exists somewhere. Unless there is some such facility, the important questions mentioned above will continue to be answered on the basis of who has the loudest voice or the best anecdote.¹⁵⁸

Open data about the judicial system is only in its infancy; in many cases the data is either not open or is never collected due to a lack of resources or political will. But that appears to be changing. As some judicial institutions successfully open their data to outside scrutiny, more may follow.

D. Uncovering Human Rights Abuses

Open data offers benefits for human rights advocacy and progressive policymaking. As a matter of principle but also to enable innovative problem-solving, the United Nations and the World Bank have made commitments to open their own data, which includes information relevant to promoting both human rights and development.¹⁵⁹ And, as discussed above,¹⁶⁰ the interplay between the United Nations' SDGs and open data provides an opportunity for data to help with measuring progress toward the SDGs, to continue exposing shortcomings of governments in realizing development-related human rights, and to inform evidence-based policymaking in response.

Beyond sustainable development, governments throughout the world have made relevant sources of data available pursuant to open government mandates that are relevant for the realization and the protection of human rights. One example worth considering is the potential impact of open data on the right to safe working conditions.¹⁶¹ In the United States, the

158. Peter Zimroth, *Note on the Fall Conference: Consumer Class Actions*, CENTER ON CIVIL JUSTICE AT NYU SCHOOL OF LAW (Dec 3, 2014), <http://www.law.nyu.edu/centers/civiljustice/developmentsandtrends/december2014> [<https://perma.cc/455E-AQEM>].

159. The United Nations Open Data Catalog is available at <http://data.un.org> [<https://perma.cc/8XCR-AS54>]. The World Bank Open Data Catalog is available at <http://data.worldbank.org> [<https://perma.cc/UT7R-LU8T>]. For more information about the World Bank's decision to open its data, see Press Release, World Bank Group, World Bank Group Opens Data to All, No. 2010/350/EXT (Apr. 20, 2010), <http://www.worldbank.org/en/news/press-release/2010/04/20/world-bank-group-opens-data-to-all> [<https://perma.cc/WZ68-KAU7>].

160. See *supra* Part II.E.

161. The International Covenant on Economic, Social and Cultural Rights explicitly includes safe and working healthy conditions as part of a right to just and favorable working conditions. See International Covenant on Economic, Social and Cultural Rights (ICESCR) art. 7(b), *opened for signature* Dec. 16, 1966, 993 U.N.T.S. 3 [hereinafter ICESCR]. Note, however, that the United States has not yet ratified the ICESCR. The United States did however ratify the International Convention on the Elimination of All Forms of Racial Discrimination, which sets out an obligation of non-discrimination in realizing the right to just and favorable working conditions. See International Convention on the Elimination of All Forms of Racial Discrimination art. 5(e)(i), *opened for signature* Dec. 21, 1965, S. EXEC. DOC. No. C, 95-2 (1978),

Occupational Health and Safety Administration (OSHA) publishes workplace safety and fatality data in fully downloadable formats, enabling better monitoring and more targeted regulatory oversight.¹⁶² In this publically accessible form, OSHA's data can also be combined with crowdsourced data from workers, for example, to develop a more real-time picture of workplace safety and to test predictors of risk.

Where institutions submit voluntarily or subject to regulatory mandate, open data has obvious benefits. Underexplored, however, is the emerging benefit that open data offers for monitoring human rights abuses. Activist and watchdog groups are opening the data they have collected and combining it with public open data to create new monitoring tools. For instance, the CIRI Human Rights Data Project combines the U.S. Department of State's *Country Reports on Human Rights Practices* and annual reports from Amnesty International to offer a customizable and downloadable dataset on fifteen human rights from 202 countries from 1981-2011.¹⁶³

The Human Rights Data Analysis Group (HDRAG) generates data-driven reports on human rights violations and creates the Human Rights Atlas, an interactive human rights map.¹⁶⁴ HRDAG uses data from "individual testimonies, legal depositions, probability surveys, administrative records from morgues and cemeteries, exhumation reports, operational records from a prison, career information on military and police officers, eyewitness interviews, and official customs and immigration records."¹⁶⁵ The group notes that its "primary focus . . . is on the rigorous scientific analysis of our partners' data."¹⁶⁶ Such efforts become exponentially easier when governments make the data available.

Open data is also enabling more direct action to correct human rights and civil liberties abuses because of the collective action and crowdsourcing that become possible when everyone can see the data. In the United States, open data is allowing individuals to protect their rights to participate in "periodic elections [of] universal and equal suffrage"¹⁶⁷: DistrictBuilder, an open source project, uses census and election data to analyze the constitutionality of legislatively-created redistricting plans and to enable

660 U.N.T.S. 195.

162. See *Data and Statistics*, OCC. HEALTH & SAFETY ADMIN., <https://www.osha.gov/oshstats> [<https://perma.cc/K2Z9-YU2H>].

163. FAQ, CIRI HUMAN RIGHTS PROJECT, <http://www.humanrightsdata.com/p/faq.html> [<https://perma.cc/4QN7-RN8H>].

164. *About Us*, HUMAN RIGHTS DATA ANALYSIS GROUP, <https://hrdag.org/aboutus> [<https://perma.cc/T3XA-3PCR>].

165. *Id.*

166. *Id.*

167. Article 25(b) of the International Covenant on Civil and Political Rights guarantees to each citizen of state parties the right and the opportunity "to vote and to be elected at genuine periodic elections which shall be by universal and equal suffrage and shall be held by secret ballot, guaranteeing the free expression of the will of the electors." International Covenant on Civil and Political Rights art. 25(b), *opened for signature* Dec. 19, 1966, S. EXEC. DOC. No. E, 95-2, 999 U.N.T.S. 171 [hereinafter ICCPR].

communities to transparently draw their own district lines.¹⁶⁸

In Ukraine, open data has helped to protect the human right to adequate housing.¹⁶⁹ Since temperatures often dip below zero degrees Fahrenheit in Ukraine, adequate heat is critical. When residents of one municipal apartment building were left freezing in their homes after contracted repair work to improve insulation never materialized even though the contract had been paid, open data provided an avenue for a collaborative solution. Using the ProZorro spending website to identify the problem, local group Anticorruption Headquarter pressured the local government and the responsible contractor to fix the problem.¹⁷⁰

In some cases, of course, governments fail to provide such open data. When this happens, activists have often created or collected that data themselves, in some cases challenging pre-existing citizen-government power dynamics. For example, in many places across the globe, so-called “citizen scientists” are gathering data to complement and fill in gaps in administrative data—some intentional, some accidental. One such example came during the 2011 Fukushima Daiichi Nuclear Disaster in Japan. Distrustful of government-published information,¹⁷¹ citizens began collecting data of their own using handheld Geiger counters, which was compiled, monitored, and openly shared through a project known as Safecast.¹⁷² Safecast, which now coordinates citizen engagement to monitor nuclear safety and air quality around the world, describes its mission as “provid[ing] citizens worldwide with the tools they need to inform themselves by gathering and sharing accurate environmental data in an open and participatory fashion.”¹⁷³ Safecast publishes all the data it collects as open data.¹⁷⁴

168. *Redistricting*, U.S. ELECTION PROJECT, <http://www.electproject.org/redistricting> [<https://perma.cc/U2EU-AF49>]. The principal investigators on DistrictBuilder, which has won numerous awards for data innovation and software in the public interest, are Micah Altman and Michael McDonald. *Id.*

169. ICESCR Article 11 protects the right to adequate housing. ICESCR, *supra* note 161, art. 11. As explained later by the Committee on Economic, Social and Cultural Rights, “[a]n adequate house must contain certain facilities essential for health, security, comfort and nutrition. All beneficiaries of the right to adequate housing should have sustainable access to natural and common resources, safe drinking water, energy for cooking, heating and lighting, sanitation and washing facilities, means of food storage, refuse disposal, site drainage and emergency services.” Comm. on Econ., Soc. & Cultural Rights, General Comment No. 4: The Right to Adequate Housing, ¶ 8(b), U.N. Doc. No. E/1992/23, annex III (Dec. 13, 1991) (emphasis added).

170. See Heyman, *supra* note 112.

171. See Azby Brown et al., *Safecast: Successful Citizen-Science for Radiation Measurement and Communication after Fukushima*, 36 J. RADIOLOGICAL PROTECTION 582, 583-84 (2016).

172. *Id.*

173. *About*, SAFECAST, <http://blog.safecast.org/about> [<https://perma.cc/XN7C-JW9L>].

174. *Id.*

E. Reducing Abuse and Enhancing the Impact of Development Assistance

Open data is facilitating projects that enable scrutiny of aid, both ensuring that it does not get diverted away from its original purposes and helping to expose its political dimensions. Open Development Cambodia, for example, uses data from governmental and non-governmental sources to provide insight into development expenditures and their impacts.¹⁷⁵ Another website, AidData, tracks international development funding.¹⁷⁶ Their analysis of the data has demonstrated that China provides more foreign aid to those African countries that support their votes in the United Nations General Assembly than those countries which do not.¹⁷⁷

As the shift toward big and open data-driven projects continues, the inherent opportunity for collaborative approaches appears to have captured a public-interest minded audience. As observed in the Open Data Barometer's most recent Global Report, "once the preserve of academics and statisticians, data has become a development cause embraced by everyone from grassroots activists to the U.N. Secretary-General. There's now a clear understanding that we need robust data to drive democracy and development—and a lot of it."¹⁷⁸ The benefits of this data go far beyond pursuing and monitoring the SDG targets.¹⁷⁹

Pursuing such promise and possibilities, the ecosystem of open data is populated by an array of stakeholders who are willing to invest in the co-creation of government policies and services, to drive "innovation that can improve education and health care, create new businesses, and stimulate scientific progress."¹⁸⁰ One example of this is Open Data for Development (OD4D), a global partnership of organizations committed to advancing the creation of "locally-driven and sustainable open data ecosystems."¹⁸¹ Through a series of initiatives that focus attention on the importance of open data for improving people's lives, the OD4D seems to scale "open data approaches that work, improve transparency and accountability, service delivery, and the well-being of the poorest and most marginalized."¹⁸²

175. *Background*, OPEN DEV. CAMBODIA, <https://opendevelopmentcambodia.net/about/background> [https://perma.cc/KXM5-9MQ5].

176. *See Our Story*, AIDDATA, <http://aiddata.org/our-story> [https://perma.cc/Y927-C3H5].

177. *Chinese Aid to Africa Looks a Lot Like Western Aid to Africa, New Research Finds*, AIDDATA: BLOG (Oct. 15, 2015), <http://aiddata.org/blog/chinese-aid-to-africa-looks-a-lot-like-western-aid-to-africa-new-research-finds> [https://perma.cc/F6ZG-XFBH]; *see also Diplomacy and Aid in Africa*, THE ECONOMIST (Apr. 14, 2016), <http://www.economist.com/blogs/graphicdetail/2016/04/daily-chart-10> [https://perma.cc/2J4H-XCUW].

178. *Open Data Barometer: ODB Global Report*, *supra* note 77, at 6.

179. *Id.*

180. *Id.*

181. OPEN DATA FOR DEV., <http://od4d.net> [https://perma.cc/2CHX-USKY].

182. *Annual Report 2015*, OPEN DATA FOR DEV. 11 (2015), <http://od4d.com/wp->

OD4D is just one of many efforts emerging to help governments open their data and enable civil society to create the capacity to collaborate in using data to solve problems. Going forward, knowing where and how to invest resources will depend upon having a clear understanding of the social, economic, and political impacts of open data on development in different contexts and a deeper awareness of the skills needed to make good use of data to achieve institutional improvements.

V. OPEN DATA AND FOIA: COMPLEMENT NOT REPLACEMENT

The explosion of open data, coupled with the development of technologies to disseminate and understand it, is cause for optimism. That said, open data cannot be the sole tool for sparking government information-sharing. While substantial, the current inventory of open datasets still represents only a small share of important government data. For instance, the World Wide Web Foundation concludes that, of 1,380 datasets they identify as priority datasets to publish, almost 90% are still closed, while much of what is open is of poor quality.¹⁸³

Of course, open data's limitations extend beyond just the slow pace of implementation. Gaps still exist in the legal regimes governing both open data and FOIA. Nevertheless, I submit that an approach that blends the two—combining the collaborative with the adversarial, the ex post with the ex ante, proactive publication with demand-driven release—creates a more effective, complementary system for information disclosure and, ultimately, results in more effective governing.¹⁸⁴

This section opens by outlining several of open data's most pressing shortcomings before proposing five practical steps to harness the best of both regimes while also addressing their limitations.

A. Open Data's Potential Shortcomings

Challenge 1: Political Will

First, and most obviously, the myriad open data success stories documented throughout this Essay depend upon the political will to be

content/uploads/2016/06/OD4D_annual_report_2015.pdf [https://perma.cc/9J4U-G9QR].

183. *Open Data Barometer: ODB Global Report*, *supra* note 77, at 7.

184. Such a complementary system of disclosures is particularly valuable where governments identify data as particularly sensitive and are therefore not likely to have the interest in making them regularly available. Consider, for instance, the contents of the Torture Database (<https://thetorturedatabase/>), an American Civil Liberties Union repository of FOIA-obtained documents related to post-9/11 policies of enhanced interrogation techniques, rendition, and detention practices. See *About*, ACLU: THE TORTURE DATABASE, <https://www.thetorturedatabase.org/about> [https://perma.cc/NZF9-RSZT]. This type of data would likely not have been willingly released—and was not willingly responded to through FOIA—by the government. See *id.* (noting the role of litigation in processing this FOIA request).

transparent and collaborative. Governments of all political stripes refuse to disclose data when they should. The data on Data.gov is often not the contentious nuclear secrets, budget models, or national security information most in demand by journalists, activists, and researchers. Indeed, there is a looming risk that governments, seeking recognition for their proactive disclosure, will only post data which is expedient and uncontroversial to publish—a practice increasingly referred to as “open-washing.”¹⁸⁵

Further complicating this issue is that political will is not uniform. A government that champions open data in some areas may wish not to disclose embarrassing or contentious information in others. For example, the Obama Administration did not produce its budget models with a transparent methodology. “One will look in vain,” writes David Schoenbrod, “in the 2015 Financial Report for a succinct statement of why it projects a much smaller gap between spending and receipts for the 75-year period than did the 2014 report.”¹⁸⁶

Nor is political will static. Especially as political administrations change, there is a risk that changes in political will can undermine existing accomplishments in open data. When administrations change from leaders who make open data hallmarks of their administrations to ones who do not, public institutions may reflect their administrations’ priorities and stop identifying and posting open data. This transition occurred in 2016 in both the United States and United Kingdom, as President Trump and Prime Minister May replaced President Obama and Prime Minister Cameron, respectively. Such changes can have obvious and concrete results: President Trump’s administration has put funding for the census in jeopardy,¹⁸⁷ employs a key fiscal official who has voted to defund the long-form American Community Survey,¹⁸⁸ and has threatened to eliminate some of NASA’s climate science missions.¹⁸⁹ Individuals and scientists fear that the Trump Administration ultimately may choose not to collect and publish important data at all.¹⁹⁰ Even data that seems less important on its face, like

185. See, e.g., Lindsay Ferris & Julia Keserü, *We Should Demand More from ‘Open’ Than Just Data: Thoughts on the International Open Data Conference*, SUNLIGHT FOUND.: BLOG (June 3, 2015), <https://sunlightfoundation.com/2015/06/03/we-should-demand-more-from-open-than-just-data-thoughts-on-the-international-open-data-conference> [<https://perma.cc/36F7-3KCH>].

186. DAVID SCHOENBROD, D.C. CONFIDENTIAL: INSIDE THE FIVE TRICKS OF WASHINGTON 106 (2017).

187. See *How Science Fares in the U.S. Budget Deal*, SCIENCE (May 1, 2017), <http://www.sciencemag.org/news/2017/05/how-science-fares-us-budget-deal> [<https://perma.cc/G52W-77LG>] (describing how the budget insufficiently allocates funds for 2020 Census preparations).

188. Jeffrey Mervis, *Scientists Fear Pending Attack on Federal Statistics Collection*, SCIENCE (Jan. 3, 2017), <http://www.sciencemag.org/news/2017/01/scientists-fear-pending-attack-federal-statistics-collection> [<https://perma.cc/G3S3-7A5N>] (discussing current Office of Management and Budget Director Mick Mulvaney’s 2012 vote to defund the long-form American Community Survey).

189. Henry Fountain, *Scientists Fear Climate Data Gap as Trump Aims at Satellites*, N.Y. TIMES (Apr. 10, 2017), <https://www.nytimes.com/2017/04/10/climate/trump-nasa-satellites-global-warming-data.html> [<https://perma.cc/2SHQ-Q6A5>].

190. See, e.g., *Politics Podcast: Data Under Trump*, FIVETHIRTYEIGHT (Jan. 2, 2017)

meeting logs and websites supporting open data, have already disappeared.¹⁹¹ Sadly, these practices will be subject to the vagaries of politics, and will, of necessity, spark citizens to try to back up open data before it disappears.¹⁹²

For some politicians, the lack of political will around open data may not be ideological or overtly ill-intentioned. Instead, a lack of political will may arise from understandable practical concerns about the financial or human capital necessary to pursue such open data objectives. Indeed, a lack of data analytic capacity is a frequently cited complaint as to why governments are not able to collect, clean, publish, or use more data. Although there is work to be done to ensure that politicians around the world understand the value of open data, assuaging these concerns will be complicated.

Challenge 2: The Lack of a Right of Action

Political will, which is a grave concern in relation to open data, is, of course, less of a concern under FOIA, as FOIA does not provide governments with a choice about disclosure (though governments may delay or fight rather than disclose). Under FOIA, governments legally must disclose requested information (possibly with acceptable redactions), provided it does not fall within pre-defined exemptions.¹⁹³ In contrast, where the political will to publish a given dataset is absent, open data laws lack a private right of action to compel disclosure. The inability to compel disclosure can cause extreme difficulty and frustration for citizens seeking access to important information.

(FiveThirtyEight colleagues in different fields discuss government data and transparency under the Trump administration.); Mona Chalabi, *Statisticians Fear Trump White House Will Manipulate Figures to Fit Narratives*, THE GUARDIAN (Jan. 30, 2017), <https://www.theguardian.com/us-news/2017/jan/30/statistics-trump-administration-numbers-manipulation> [https://perma.cc/8EF3-RN9Q].

191. See Julie Hirschfeld Davis, *White House To Keep its Visitor Logs Secret*, N.Y. TIMES (Apr. 14, 2017), <https://www.nytimes.com/2017/04/14/us/politics/visitor-log-white-house-trump.html> [https://perma.cc/76B6-YRFE] (explaining the Trump administration's revisions to the Obama administration policy of disclosing who visits the White House); Tracie Mauriello, *Government Watchdogs Criticize Trump's Removal of Open Data Sets*, GOV'T TECH. (Feb. 22, 2017), <http://www.govtech.com/data/Government-Watchdogs-Criticize-Trumps-Removal-of-Open-Data-Sets.html> [https://perma.cc/6Z88-WEQS] (discussing the Trump Administration's changes to <http://open.whitehouse.gov>).

192. In the run-up to President Trump's inauguration, many groups raced to backup open data, such as environmental data, lest it be taken down. See Brady Dennis, *Scientists Are Frantically Copying U.S. Climate Data, Fearing It Might Vanish Under Trump*, WASH. POST (Dec. 13, 2016), <https://www.washingtonpost.com/news/energy-environment/wp/2016/12/13/scientists-are-frantically-copying-u-s-climate-data-fearing-it-might-vanish-under-trump> [https://perma.cc/7YSL-YHQZ]. One group engaged in this work is the Environmental Data and Governance Initiative (EDGI), which is responding to future environmental deregulation by collecting and archiving environmental datasets, creating usable non-governmental data access, and preserving internet-based environmental policy and program records. ENVTL. DATA & GOVERNANCE INITIATIVE, <https://envirodatagov.org> [https://perma.cc/2AJQ-A2PP]. EDGI hosts events that bring together coders, archivists and people with environmental expertise. *Id.*

193. GINSBERG, *supra* note 19, at 2.

Although the reintroduced OPEN Government Data Act would compel agencies to inventory and publish all government data in machine-readable formats (the legislation's exemptions are the same as those under FOIA),¹⁹⁴ this legislation fails where much of open data policy does: it lacks teeth in the form of a legal right of action.

The legal right of action to sue (or threaten to sue) under FOIA for information disclosure is essential to gain access to data of public import. For example, in 2013, transparency activist Carl Malamud began coordinating an effort to use FOIA to force the IRS to publish nonprofit tax returns as open data.¹⁹⁵ Malamud had to use FOIA successfully to request nine nonprofit tax returns from the IRS because the agency would not make the returns available in digital form.¹⁹⁶ Although disclosure of nonprofit returns is required by law and the filers submitted those returns electronically, the IRS wanted to send Malamud image files of the returns. The IRS typically took electronically-filed returns, printed them out, scanned them back in, and sold DVDs with the image files.¹⁹⁷ Because of his successful suit and campaign, the IRS not only turned over Malamud's nine requested returns in a digitally readable format¹⁹⁸ but soon after began to make all electronically filed nonprofit tax returns, which represents about 60% of those filed since 2011, digitally downloadable as open data.¹⁹⁹

As the Malamud example makes clear, when collaboration fails, litigation can sometimes help to spur publication and give momentum to reformers inside government looking for the push needed to overcome resistance; the IRS now actively publishes Form 990 data. Open data researchers have been quick to make wise use of this development. For example, in May 2017, the Aspen Institute's Program on Philanthropy and Social Innovation hosted a "Datathon," bringing together nonprofit researchers and data scientists to "begin the labor-intensive process of cleaning and converting Form 990 data into more accessible, public spreadsheets."²⁰⁰ In light of developments like the "Datathon," there is little question that a legal right of action complements open data-style

194. OPEN Government Data Act, S. 760, 115th Cong. (2017).

195. Suzanne Perry, *IRS Plans To Begin Releasing Electronic Nonprofit Tax Forms Next Year*, CHRON. OF PHILANTHROPY (June 30, 2015), <https://www.philanthropy.com/article/IRS-Plans-to-Begin-Releasing/231265> [<https://perma.cc/67HY-Q7M7>].

196. *Id.*

197. See BETH SIMONE NOVECK & DANIEL GOROFF, *Information for Impact: Liberating Non-Profit Data*, ASPEN INSTIT. 15 (2d ed. 2013), https://assets.aspeninstitute.org/content/uploads/files/content/docs/pubs/Information_for_Impact_Report_FINAL_REPORT_9-26-13.pdf [<https://perma.cc/6ZFU-LDCX>].

198. Suzanne Perry, *IRS Releases Tax Forms of 9 Nonprofits in Computer-Readable Format*, CHRON. OF PHILANTHROPY (June 8, 2015), <https://www.philanthropy.com/article/IRS-Releases-Tax-Forms-of-9/230759> [<https://perma.cc/BGL5-LTJE>].

199. *IRS Makes Electronically Filed Form 990 Data Available in New Format*, INTERNAL REVENUE SERV., IR-2016-87 (June 16, 2016), <https://www.irs.gov/uac/newsroom/irs-makes-electronically-filed-form-990-data-available-in-new-format> [<https://perma.cc/2BKN-MN7H>].

200. Stephanie Román, *The Aspen Institute's Program on Philanthropy and Social Innovation (PSI) Hosts Nonprofit Datathon*, ASPEN INSTIT. (June 6, 2017), <https://www.aspeninstitute.org/blog-posts/aspen-institutes-program-philanthropy-social-innovation-psi-hosts-nonprofit-datathon> [<https://perma.cc/HM8Q-VGK4>].

collaboration.

Challenge 3: Data Invisibles

Of course, where data is not collected at all, especially about socially and economically disadvantaged groups, or is incomplete, inconsistent and unreliable, FOIA and open data are both of little use. Thus, there is need to adopt every approach that can help address the ethical implications for the “digital invisibles” – people on whom data are not collected. A group of Canadian researchers recently found that people who lack representation in the big-data world may be subject to misguided interventions and biased policy.²⁰¹

U.S. federal crime data is a good example of how certain data is collected while other relevant data is ignored. Although the FBI collects, publishes, and makes available for downloading the Uniform Crime Reports’ estimated monthly aggregates of instances of eight major crimes (murder, rape, assault, robbery, arson, burglary, larceny-theft and motor vehicle theft),²⁰² we have no similar data store for white collar crime. Crime in the United States is “arguably, at least as much about corporate fraud as about armed robbery, harassment via the Internet as about breaking and entering, and endangering health through environmental pollutants as about assaults and muggings.”²⁰³ Despite this, early twenty-first century information infrastructure collects only data about “so-called ‘street crime’ – violent crime and some types of property crime – to the general exclusion of non-street crime.”²⁰⁴ At the national level in the United States, for example, despite a statutory requirement to collect it, data relating to shootings and deaths in police custody is practically non-existent.²⁰⁵ The existence of data changes policy priorities.

The failure to collect data is but one of many reasons why public institutions are facing challenges with uncovering the rich insights within their own data. The lack of data analytical talent within the public sector also serves to render disadvantaged persons and groups invisible.²⁰⁶

201. Justin Longo et al., *Technology Use, Exposure to Natural Hazards, and Being Digitally Invisible: Implications for Policy Analytics*, 9 POL’Y & INTERNET 76, 99-100 (2017).

202. Fed. Bur. Investigation, *UCR Offense Definitions*, U.S. DEP’T OF JUSTICE, <https://www.ucrdatatool.gov/offenses.cfm> [<https://perma.cc/Z3QF-RNX6>].

203. MODERNIZING CRIME STATISTICS: REPORT 1: DEFINING AND CLASSIFYING CRIME 10 (Janet L. Lauritsen & Daniel L. Cork eds., 2016).

204. *Id.*

205. The Death in Custody Reporting Act, 42 U.S.C. § 13727a (2014), requires collection and reporting of data about how many people die when detained or incarcerated, but, as of late 2016, the data collection practices under that statute had still not been implemented. See Office of Pub. Affairs, *Justice Department Outlines Plan to Enable Nationwide Collection of Use of Force Data*, U.S. DEP’T OF JUSTICE (Oct. 13, 2016), <https://www.justice.gov/opa/pr/justice-department-outlines-plan-enable-nationwide-collection-use-force-data> [<https://perma.cc/K567-G9CL>].

206. See Beth Simone Noveck, *Another NHS Crisis Looms – An Inability To Analyse Data*, THE GUARDIAN (Feb. 8, 2017), <https://www.theguardian.com/science/political-science/2017/feb/08/another-nhs-crisis-looms-an-inability-to-analyse-data>

Although outside the scope of this Essay to discuss further, the lack of data analytical capacity remains one of the most fundamental challenges impeding transparency and evidence-based governing. Neither open data nor FOIA law address this institutional challenge.

Challenge 4: Prioritization

Open data, at least in theory, requires inventorying *all* data and creates an opportunity for reasoned debate between the public and the agency about what to publish, with what frequency, and in what formats. But because open data is often the creature of executive action and not of legislation, open data is not always a systematic process with a clear definition of high-value information.

There is often no rhyme or reason behind what is defined as “high-value” information and therefore what gets published.²⁰⁷ The lack of a cognizable or sensible publication prioritization can frustrate and hinder the uptake of open data. To be sure, agencies post a lot of data on Data.gov that no one wants, knows exists, or uses.²⁰⁸

As civil servants learn about the benefits of open data, a sense of priorities may emerge. Eventually, disclosure prioritization might be mandated through legislation. However, this hope is not certain to become reality.

Thus, we must ask and try to measure what kind of discourses and politics will be produced by a transparency regime that has no lodestar. If open data is oriented toward publication of easily understood and quantifiable information that technical people can turn into consumer tools (such as transit data that becomes a “when is my bus coming?” app), will other kinds of substantive information be neglected in favor of disclosures that lead to headline-grabbing consumer tools like the College Scorecard? What kind of information ecology results if the complex information that forms the basis of government decision-making, such as budget models,

[<https://perma.cc/D4HR-BQQ5>]; Beth Simone Noveck et al., *Smarter State: Smarter Health: Boosting Analytical Capacity at NHS*, GOVLAB (Feb. 2017), <http://www.thegovlab.org/static/files/publications/nhs-health.pdf>

[<https://perma.cc/B5B2-R35T>] (“The collection and analysis of data, when done responsibly and in a trusted manner, has the potential to improve treatment and drive towards value, both social and economic, in healthcare. However, the goal of using data to improve the NHS and social care is hampered by a “talent gap” — a lack of personnel with data analytical skills — that stands in the way of uncovering the rich insights expected to reside in the NHS’ own data. The NHS is not unique among public and even private sector institutions who are struggling to identify, hire and retain people with data science skills, and, above all, with the ability to apply new technological tactics to advancing the institution’s priorities.”).

207. The Open Government Directive does provide a broad definition for high-value data. See Orszag, *supra* note 23, at 7-8 (defining high-value datasets). However, this has not necessarily provided sufficient direction. See discussion and criticisms of the looseness of high-value data definitions *supra* Section I.B.

208. Even where data is, in theory, posted to Data.gov, there is often a lack of investment in making the data truly accessible and usable. What one finds on Data.gov is frequently just a link to a website describing data but no actual way to download it.

gets neglected in favor of data that feeds consumer apps?

Drawing upon the lessons of FOIA—which emphasizes disclosure of data by government about government—open data policy, too, should evolve to articulate normative guidance to agencies about what should be published online and when and how to make use of it. Although there is a certain appealing optimism to the early organic and ad hoc evolution of open data practices, there is a need to evolve beyond the unsystematic apps-over-substance nature of open data policy to focus on disclosures and their uses that lead to positive social change, advance progressive values, and lead to measurable progress as measured not by the number of datasets released but by the downstream impact on people’s lives.

Because open data enables the publication of ever larger datasets that can then be analyzed using algorithms it lends itself to projects that benefit from comparisons at scale, such as analysis of the efficiency, effectiveness, and disparate impact of how policies and services are delivered. In an example of open data being used to mitigate inequitable distribution, Transparency International and the World Wide Web Foundation have an ongoing effort to help civil society and governments use open data to identify and fight corruption, especially in procurement.²⁰⁹ Opening the entire corpus of data about food-borne illnesses, to take one final example, provides a supply of information to match the demand for better algorithms that helps a city like Chicago to better allocate its scarce restaurant inspection and enforcement resources more efficiently.²¹⁰

Moving toward a more principled approach to open data also demands focusing on outcomes rather than on inputs. In the first generation, we celebrated the act of publishing datasets—transparency for its own sake—regardless of who used them and to what end, if at all. To strengthen the normative underpinnings of open data, however, it is important to start with a clear definition of the problem to be solved, be it corruption or human rights abuses or agricultural productivity, and use open data as the means to the end rather than as an end unto itself. Therefore, efforts to publicize the calendars of cabinet secretaries or the salaries of government officials may be weak candidates for open data efforts because, absent gross malfeasance, such disclosures will not drive changes in how government operates and may sap the political will for furthering open data. Open data priorities should retain their broader orientation toward high-value problem-solving.

Thus, equally important to publishing the clean, comprehensive, and timely data needed to solve a well-defined problem is also investment in

209. For their most recent report on the matter, see *Connecting the Dots: Building the Case for Open Data To Fight Corruption*, WORLD WIDE WEB FOUND. & TRANSPARENCY INT’L (2017), http://files.transparency.org/content/download/2109/13448/file/2017_OpenDataConnectin gDots_EN.pdf [<https://perma.cc/L354-ZQ2T>].

210. See *supra* notes 41-42 and accompanying text; see also Julian Spector, *Chicago Is Predicting Food Safety Violations. Why Aren’t Other Cities?*, CITYLAB (Jan. 7, 2016), <http://www.citylab.com/cityfixer/2016/01/chicago-is-predicting-food-safety-violations-why-arent-other-cities/422511> [<https://perma.cc/XN5B-KQ9A>].

and prioritization of the collaboration and engagement necessary to build a coalition among those interested in using the data to tackle the problem. Open data encourages efficient outcomes because the proactive disclosure is often, although not always, accompanied by a plan for how to use the data and how to cultivate citizen engagement among those interested in applying the data for social good. These areas of alignment where government and civil society or industry are prepared to collaborate to address a challenge are, obviously, excellent opportunities to realize value.

B. Proposals for Enhancing Both FOIA and Open Data

FOIA and open data both emphasize disclosure of information created or collected by the government to the public, but the normative underpinnings and the mechanics differ dramatically. In contrast to FOIA's focus on disclosing the inputs of government decision-making, open data is rooted in a theory about government legitimacy stemming from outcome-oriented effectiveness. Open data substitutes a utilitarian rationale for transparency (evidence-based decision-making) in place of FOIA's deontological justification based on moral obligation. In contrast to traditional right to information theories, the open data narrative emphasizes that information should be open, not just because of the ill effects of secrecy but also for the instrumental reason that taxpayers have paid for the collection of that data.²¹¹

This is not to say that open data does not still face significant practical or ideological obstacles. Despite considerable enthusiasm for open data as a new tool of collaborative governing, far too many datasets that should be available for public scrutiny remain unavailable, and a lack of resources and skills in many locations hinders analysis of the data that is available. In addition, there exist legitimate concerns that allowing governments to focus on data to improve consumer decision-making tempts the government to deprioritize, hide, or exclude from disclosure other information. And, at times, the open data policy and its advocates naively assume an amicable collaboration between government and civil society that does not exist.²¹²

Whatever hurdles remain, the open data movement has progressed far enough worldwide—and is so entrenched in modern technological and cooperative thought—that it is not going away. Open data law is a creature of the big data era and assumes a process by which value and insight are to be gleaned from the analysis of ever larger quantities of data, rather than

211. See Janssen, *supra* note 10 (“In general, while the RTI movement uses rights-based arguments, OGD activists rely more on the technical possibilities of data and on the proprietary argument that government data belongs to the public taxpayers.”). Note, however, that this is not a pure binary; open data may also be consistent with more modern efforts around the right to information.

212. Those who would be satisfied with open data or FOIA as the highest achievement of transparency must wrestle with the fact that some of greatest advances in transparency in recent years have come about because of extra-legal disclosures like WikiLeaks, the Panama Papers, or the Snowden disclosures.

from the smoking gun hidden in a single document. Because open data has its roots in the data processing technologies of the big data era, its enabling legislation typically contains and establishes a policy framework designed to produce ever-larger quantities of data. At this point, FOIA's "look-up culture" and the mechanics of a system designed to produce small quantities of data seems wholly inadequate to quell the current fervor for open data movement. Similarly, as data openness continues to pervade societal thought, I suspect that the instinct to cooperatively build open futures through open data will lessen our emphasis on the adversarial FOIA process.

Although open data enjoys an increasingly entrenched status, blending elements of FOIA and open data would further strengthen open data and insulate it from obstacles both political and practical. By combining key rights of legal redress to get at government secrets while taking advantage of open data's collaborative and participatory dynamics, open data will enjoy a better opportunity to reach its potential for social good. Below, I share four proposals that would yield significant progress toward achieving these aims. Although U.S.-focused in their specifics, the gist of these proposals is applicable in all jurisdictions where both FOIA and open data policies are in place.

Proposal 1: Merge Open Data with FOIA Collection and Publication Mechanisms

Because FOIA and open data are two separate regimes, governments continue to develop two uncoordinated transparency processes. Instead of this bifurcated approach, governments must take steps to merge open data with FOIA collection and publication mechanisms.

To guarantee ease of access, governments should establish a single website for information requests, whether pursuant to FOIA or open data policy, and publicly post information in response to those requests. Furthermore, to meet open data requirements, agencies should not be given a choice about participation. Instead, all agencies should be required to participate in and use the new portal. As much as possible, the process should take advantage of the existing technology and infrastructure (in the United States, this is Data.gov, which is housed by the General Services Administration). Ultimately, open data and FOIA requests and publication should blend into a "one-stop shop" for requesting and searching all government information.²¹³

213. Notably, the FOIA Improvement Act passed in 2016 goes part of the way. FOIA Improvement Act of 2016, Pub. L. No. 114-185, 130 Stat. 538 (to be codified at 5 U.S.C. § 552) ("copies of all records, regardless of form or format—i) that have been released to any person under paragraph (3); and ii) (I) that because of the nature of their subject matter, the agency determines have become or are likely to become the subject of subsequent requests for substantially the same records; or (II) that have been requested 3 or more times"). It calls for setting up a single electronic portal for FOIA requests and although it is silent as to where the responses to those requests will be posted. *Id.* (to be codified at 5 U.S.C. §552(e)(4)). I submit

Proposal 2: Publish All Information Requested Pursuant to Either Freedom of Information or Open Data Law in Machine-Readable Formats

Going forward, all information requested pursuant to either FOIA or open data law or policy should be published in machine-readable formats.²¹⁴

Although in the United States the FOIA Improvement Act calls for releasing information electronically, it is silent as to the format.²¹⁵ It is possible to lessen the harmfulness of this error without the delay of legislative solutions. Without waiting for Congress, the Department of Justice, in collaboration with the FOIA Ombudsman in the National Archives and the Office of Management and Budget (or those responsible for information management in other countries), can and should issue guidance defining the FOIA electronic formats required under the law as those under the open data regime. The formats must be machine-readable and should post pointers back to the data they put up online on Data.gov or the relevant national platform. Again, the gist of this idea should be implemented in all jurisdictions.

This method would eventually bring about numerous benefits. As agencies shift to a “release-to-one-release-to-all” strategy and post information online, requestors can search for desired information prior to filing a new request. In addition to enabling greater innovation, such as data analysis and visualization, as the public takes hold of the released information, this strategy should cut down on requests—including political nuisance requests—and processing times for non-contentious information.

Proposal 3: Increase Dialogue Between Officials Responsible for FOIA and Those Responsible for Open Data

Bringing FOIA into the era of big data and achieving the goal of a unified database for requests and publications will be accelerated, as a practical matter, by increased dialogue between those government officials responsible for FOIA and those responsible for open data.

In the United States, management and responsibility of FOIA and open data processes remains complicated and scattered. The Department of Justice and the General Services Administration have responsibility for

that it would make sense to use Data.gov (with a pointer from FOIA.gov) as a one-stop shop for requesting *and* searching for information. Data.gov is already set up to point to data housed across the federal government (in only limited cases is the data stored on Data.gov servers) and it has an underlying information architecture, which would make it possible to sort and search by agency or by topic.

214. *Open Data Transition Report: An Action Plan for the Next Administration*, CTR. FOR OPEN DATA ENTERPRISE 18 (2016), <http://reports.opendataenterprise.org/transition-report.pdf> [<https://perma.cc/VFD7-DCNN>].

215. See FOIA Improvement Act of 2016, Pub. L. No. 114-185, 130 Stat. 538 (to be codified at 5 U.S.C. § 552). Note that the reintroduced OPEN Government Data Act, S. 760, 115th Cong. (2017) would help remedy this problem.

FOIA and open data processes, respectively.²¹⁶ Thus there is an array of personalities, from different backgrounds, involved. This includes those with legal training, like general counsels and FOIA officers, on the one hand, and chief information officers (CIOs), chief innovation officers, and chief data scientists, with technical training, on the other.²¹⁷

Beyond these managing bodies, expertise and experience in FOIA and open data can be found throughout the federal government. For example, Data.gov was collaboratively developed by the CIO Council with engagement and input from all federal agencies.²¹⁸ Other experts in government transparency include those staffing the Office of Government Information Services (OGIS), staff of the Office of Information Policy at the Department of Justice (which oversees FOIA policy), and the various FOIA officers embedded within government agencies.²¹⁹ The bifurcation of responsibility between lawyers and technologists is not unique to the United States but characteristic of information governance worldwide: open data is technical while freedom of information is legal.

These legal and technology communities need to collaborate and agree on data publication standards and more efficient workflows as well as on strategies for ensuring that data publication helps to achieve agencies' core missions. Chief information officers, chief data officers, chief innovation officers, and others in charge of open data should be meeting to ask and answer how FOIA processes could take advantage of the technological affordances of big data and, above all, to embrace more of the collaborative nature of open data practices.

By the same token, open data managers must work to understand the audiences for FOIA and to develop more responsive approaches. This need

216. *Frequently Asked Questions*, FOIA.GOV, <https://www.foia.gov/faq.html> [<https://perma.cc/2MFA-MPX4>] ("The Office of Information Policy at the Department of Justice is responsible for issuing government-wide guidance on the FOIA as part of its responsibilities to encourage all agencies to fully comply with both the letter and the spirit of the FOIA.") By contrast, Data.gov is hosted by the U.S. General Services Administration's Technology Transformation Service. See *About*, DATA.GOV, *supra* note 53.

217. Theo Douglas, *The Job of Data: The Chiefs in Charge of Data in the C Suite*, GOV'T. TECH. (Mar. 2017), <http://www.govtech.com/people/The-Job-of-Data-The-Chiefs-in-Charge-Data.html> [<https://perma.cc/S986-NHUH>] (describing the technology professionals in charge of open data).

218. Project Open Data, an "online collection of code, best practices, and case studies developed to help agencies adopt the framework presented in the OMB memorandum M-13-13 Open Data Policy-Managing Information as an Asset" is "managed by the Office of Management and Budget (OMB) and Office of Science and Technology Policy (OSTP), both components of the Executive Office of the President. The Federal CIO and CTO will both be actively involved, along with members of their teams." *Governance*, PROJECT OPEN DATA, <https://project-open-data.cio.gov/governance> [<https://perma.cc/4CW4-MV7Z>].

219. The Director of OGIS and the Director of the Department of Justice's Office of Information Policy are both lawyers. See *Office of Government Information Services (OGIS): Director's Bio*, U.S. NAT'L ARCHIVES & RECORDS ADMIN., <https://www.archives.gov/ogis/about-ogis/directors-bio> [<https://perma.cc/4ZAM-Z8WN>]; *Office of Information Policy: Meet the Director*, DEP'T OF JUSTICE (Feb. 14, 2017), <https://www.justice.gov/oip/meet-director> [<https://perma.cc/F4YZ-T5SC>]. For information about compliance roles and responsibilities of FOIA Officers, see FOIA Improvement Act of 2016, Pub. L. No. 114-185, 130 Stat. 538 (to be codified at 5 U.S.C. § 552), § 552(j)(3).

for more internal dialogue is mirrored by the need for more conversation between the transparency and accountability interest groups,²²⁰ which have traditionally stewarded the implementation of the Freedom of Information Act and other sunshine provisions, and the open data groups,²²¹ which often have a more technological bent and whose tactics focus on engendering collaboration rather than litigation. Evolving our legal and policy framework for public information collection and publication will benefit from lawyers, technologists, data scientists, and policymakers talking more to one another and to the various constituencies that depend upon both open data and FOIA.

Proposal 4: Create and Store Data Digitally in a Searchable Cloud

In this world of big data, the bigger the quantity of information, the easier it becomes to sort and compare within and across organizations and over time to generate insights. Therefore, governments should work toward creating and storing data digitally in a searchable cloud.

With government data more comprehensively in the cloud,²²² it will be possible to run searches broadly about the workings of government, the economy, society, and the environment. One could, for example, search all public-sector email to see the topics decision-makers discuss over time. Furthermore, when it becomes possible to search all contracts or grants data across agencies and across levels of government, it should be possible to obtain a much more accurate picture of what government does and to measure its impact, applying algorithms to spot patterns of fraud, waste and corruption. Such uses of the cloud could render outdated prior notions of ferreting out a single document or posting datasets.

Open data has yet to reach its potential in this area. For example, if one wanted to conduct research on military spending by monitoring contract solicitations from a defense agency, one could do so by reading solicitations and contracts published openly online; however, downloading and working with that data is problematic today. Analysis that aims to spot trends and patterns in government spending might require going to the website repeatedly or running a computer program to draw down a massive data file and compare it to yesterday's or last year's entries.²²³ Completing a more detailed analysis (perhaps how eventual federal outlays map to the timing of original contract solicitations) would involve yet an

220. Such groups in the United States include the Program on Government Oversight (POGO) and the National Freedom of Information Coalition.

221. Such groups include the Omidyar Network, World Wide Web Foundation, Data Coalition, and Open Data Institute.

222. Of course, moving government information to the cloud will require similar attention to that paid today to publishing open data to ensure that personally-identifiable information and classified information do not inadvertently get published.

223. For an example utilizing such data, see Michael Z. Gill, *The Economic Benefits of Conflict? Estimating Defense Firm Responses to Major Events in U.S. Foreign Policy* (Aug. 2016) https://dl.dropboxusercontent.com/u/2042671/DoD_Fall2016.pdf [<https://perma.cc/PBV6-6BDL>].

additional layer of merging records across unconnected databases, such as joining unstructured contract solicitation data with the requisite procurement data.

In such a case, we would say that the data is technically “open,” but, as a practical matter, answering the question of “who is government contracting with, when, and to what degree?” is very time-consuming and requires considerable computing time, technique, and data storage available only to a select few with the time, talent, and curiosity to do the analysis. Even those inside government, who potentially have the greatest need for such analysis, may or may not have the necessary resources to run painstaking manual queries. Our public institutions too often lack the infrastructure—legal, technical, and human capacities—to support such robust and data-driven policymaking.

Often open data is, even when collected and published, still far less open than it needs to be in order to be usable. Open data, as it currently stands, is only an interim step on the pathway from a paper-based FOIA world to a future where comprehensive public information is produced and stored digitally in real-time formats that enable agile and empirical social science. Accordingly, making open data realize its potential for collaboration requires a better approach, namely the storing of comprehensive, clean, and searchable government data in a publicly accessible cloud.

Until we get there, we need to keep both FOIA and open data in our arsenal of transparency tools, using one to protect against secrecy and data-hoarding and the other to promote active, participatory citizenship and collaboration between the state and its citizens.

VI. CONCLUSION

The explosion of newly available data, coupled with mounting evidence that data catalyzes productive partnerships between government and civil society, suggests that open data will continue to grow as a tool for governing. If the trend continues, open data will lead to new empirically-informed ways of holding government and others accountable, spurring consumer choice and expanding the range of approaches to tackling human rights and development challenges.

In practice, open data promotes broad-scale transparency; simplifies the disclosure process; requires publication in reusable and computable formats; focuses on disclosure of information collected by government as regulator and researcher, rather than exclusively on data created by government about its own workings; and, above all, gets both more “eyeballs” and machines looking at the data to spot problems, identify patterns, devise solutions, and act. At the same time, as promising as open data may be, it is important to dispel any techno-utopian strain in the open data narrative that suggests that, just given enough data, all problems are solvable. FOIA indispensably *complements* open data by providing a legal right of action to compel disclosure, by suggesting the kinds of data to

prioritize releasing, and by disclosing who is using what data and how. Open data, which still depends on the political will to publish information, will be strengthened when, like FOIA, it can compel disclosure, especially in reusable formats.

But the most significant impact of open data in the long run may stem less from immediate problems tackled than from the way open data fosters more active citizenship and more collaborative democratic institutions. Open data transforms transparency policy from a means of monitoring government ex post to a mechanism for getting the public to participate actively in improving societal outcomes. By eschewing the adversarial in favor of a collaborative approach to transparency, open data reflects a radically different transparency narrative and, ultimately, a different theory of democracy whereby citizen participants collaborate on designing and building solutions to important problems together with public institutions. This collaborative model enables governments to draw directly on the collective expertise of the population in developing new and creative approaches and affords the public new opportunities to participate in the creation of a better future.

