

Southern Methodist University

SMU Scholar

Faculty Journal Articles and Book Chapters

Faculty Scholarship

2018

Regulation by Database

Nathan Cortez

Southern Methodist University, Dedman School of Law

Recommended Citation

Nathan Cortez, Regulation by Database, 89 U. Colo. L. Rev. 1 (2018)

This document is brought to you for free and open access by the Faculty Scholarship at SMU Scholar. It has been accepted for inclusion in Faculty Journal Articles and Book Chapters by an authorized administrator of SMU Scholar. For more information, please visit <http://digitalrepository.smu.edu>.

UNIVERSITY OF
COLORADO LAW REVIEW

Volume 89, Issue 1

2018

REGULATION BY DATABASE

NATHAN CORTEZ*

The federal government currently publishes 196,284 searchable databases online, a number of which include information about private parties that is negative or unflattering in some way. Federal agencies increasingly publish adverse data not just to inform the public or promote transparency, but to pursue regulatory ends—to change the underlying behavior being reported. Such “regulation by database” has become a preferred method of regulation in recent years, despite scant attention from policymakers, courts, or scholars on its appropriate uses and safeguards.

This Article evaluates the aspirations and burdens of regulation by database. Based on case studies of six important data sets (published by the CFPB, CPSC, EPA,

* Callejo Endowed Professor of Law, Gerald J. Ford Research Fellow, and Associate Dean for Research, SMU Dedman School of Law. I am grateful to Nick Bagley, Glenn Cohen, James Coleman, Mark Fenster, Chris Jenks, Jeff Kahn, Kristin Madison, Frank Pasquale, and Bill Sage for commenting on earlier drafts. I’m also grateful to participants at Harvard Law School’s Health Law Workshop, the University of Texas Law School, Gonzaga Law School, Washington & Lee Law School’s Reputation Roundtable, and the SMU Law Faculty Forum for commenting on earlier drafts. I thank Elise LeGros and Victoria Vish for their characteristically excellent research assistance. This Article builds on my recent work for the Administrative Conference of the United States (ACUS), which in June 2016 adopted recommendations for agencies publishing consumer complaint databases. See ACUS, Adoption of Recommendations, 81 Fed. Reg. 40,259 (June 21, 2016); ADMIN. CONFERENCE OF THE U.S., *Agency Information Dissemination in the Internet Era*, <https://www.acus.gov/research-projects/agency-information-dissemination-internet-era> (last visited July 16, 2017) [<http://perma.cc/HP2B-BWHY>].

FEC, FDA, and Medicare), the Article proposes what I call “Good Government Data Practices” to ensure that databases are reliable, useful, and fair. More optimal data disclosures require careful design choices that consider both data inputs and outputs, including how to gather and process data, how to characterize them, and how to present them. The article envisions a decidedly modern role for government agencies as data “stewards” rather than as mere publishers or repositories.

Agency databases have proliferated on the belief that markets, regulation, and even democracy all require transparency, that sunlight is the best disinfectant. But as transparency has moved online—becoming more pervasive, more powerful, and more burdened with regulatory dimensions—we also must recognize that sunlight can blind or even burn. It is in this spirit that I call for policymakers to embrace the government’s role as a data “steward,” a sentinel that helps maximize the quality of data inputs and outputs via tailored procedures. The more reliable government data are, the more they can enlighten us and perhaps even deter unwanted behavior.

INTRODUCTION.....	3
I. FROM OPEN GOVERNMENT TO OPEN INDUSTRY	10
A. Publishing “Secret Laws”	11
B. The APA and Notice	12
C. FOIA and Transparency by Request.....	14
D. The Internet Era	15
II. DATA AND ITS ASPIRATIONS	21
A. Market Ideals.....	21
B. Regulatory Ideals.....	23
C. Democratic Ideals	27
III. DATA AND ITS DISCONTENTS.....	28
IV. AGENCY DATABASES.....	37
A. The FEC’s Campaign Finance Data.....	37
B. The EPA’s Toxic Release Data	43
C. The CFPB’s Consumer Complaint Data.....	47
D. The CPSC’s Product Safety Data.....	52
E. Medicare’s Quality Data	55
F. The FDA’s Adverse Event Data.....	64
V. DESIGNING FOR OPTIMAL DISCLOSURE	69

2018]	REGULATION BY DATABASE	3
	A. <i>Gathering and Processing the Data</i>	71
	1. Pre-Publication Procedures	72
	2. Post-Publication Procedures for Contested Data.....	73
	3. Considering Costs.....	77
	4. Administrative Law Dimensions.....	78
	B. <i>Characterizing the Data</i>	79
	1. Identifying Sources.....	79
	2. Explaining Context, Limitations	80
	C. <i>Presenting the Data</i>	82
	1. Raw or Polished?	82
	2. Big Data or Small Data?	85
	3. Intermediaries and Collaborative Data	88
	CONCLUSION	91

Publicity is justly commended as a remedy for social and industrial diseases. Sunlight is said to be the best of disinfectants; electric light the most efficient policeman.

— *Louis D. Brandeis*¹

Perhaps we should blame it on Brandeis. As is so often the case, the perfect turn of phrase often takes on a life of its own, rendering more difficult the likelihood of careful and balanced analysis of the topic to which the phrase, like a barnacle, has become attached.

— *Frederick Schauer*²

INTRODUCTION

The federal government publishes tens of thousands of searchable online databases from hundreds of sources. The site Data.gov includes 196,284 unique data sets—roughly 150,000 from the federal government, with the remainder from sub-

1. LOUIS D. BRANDEIS, *OTHER PEOPLE'S MONEY AND HOW THE BANKERS USE IT* 92 (Thoemmes Press 2003) (1914).

2. Frederick Schauer, *The Mixed Blessings of Financial Transparency*, 31 *YALE J. REG.* 809, 809 (2014).

federal and nongovernmental sources.³ Some of these data sets include information about private parties that is negative, unflattering, or adverse in some way. For example, users can search for consumer products that may have caused injuries; for drugs that may have caused side effects; for lenders that may have treated customers unfairly; for hospitals with higher-than-average mortality rates; for airlines that lose the most luggage; for lobbyists that contribute to federal candidates; or for nearby facilities that discharge toxic chemicals.

Federal agencies often publish these data not just to inform the public or promote government transparency, but also to pursue “regulatory” aims—to influence the underlying behavior being reported. “Regulation by revelation”⁴ is not at all new, of course, with a lineage stretching back decades.⁵ By now, disclosure has been so frequently used as a tool to discourage certain conduct that it is easy to take for granted. Though it is impossible to document all regulatory frameworks that rely on disclosure, even a partial list shows how ubiquitous it has become. We now rely on disclosure to regulate food nutrition,⁶ fuel economy,⁷ hospital quality,⁸ mortgages,⁹ securities,¹⁰ sex

3. *Data Catalog*, DATA.GOV, <https://catalog.data.gov/dataset> (last visited July 16, 2017) [<http://perma.cc/6R8L-8VC4>] (see “Organization Types” in the left sidebar). Data.gov currently includes data from 166 federal agencies and subagencies. See *Federal Agency Participation*, DATA.GOV, <http://www.data.gov/metrics> (last visited July 16, 2017) [<https://perma.cc/WTJ6-RJGX>]. Note that the tally of data sets on Data.gov, including the disappearance of thousands, has been tracked closely during the transition to the new Trump administration. See Juliet Eilperin, *Under Trump, Inconvenient Data is Being Sidelined*, WASH. POST (May 14, 2017), https://www.washingtonpost.com/politics/under-trump-inconvenient-data-is-being-sidelined/2017/05/14/3ae22c28-3106-11e7-8674-437ddb6e813e_story.html [<https://perma.cc/MG8R-UJXC>].

4. Daniel C. Esty, *Environmental Protection in the Information Age*, 79 N.Y.U. L. REV. 115, 126 (2004).

5. For one of the first cross-disciplinary writings on regulation via disclosure, see William M. Sage, *Regulating Through Information: Disclosure Laws and American Health Care*, 99 COLUM. L. REV. 1701 (1999).

6. Nutritional Labeling Education Act of 1990, 21 U.S.C. § 343 (2010).

7. See *Download Fuel Economy Data*, U.S. DEP’T OF ENERGY, <http://www.fueleconomy.gov/feg/download.shtml> (last visited July 16, 2017) [<http://perma.cc/5B48-BKZE>] (providing data back to 1978).

8. See discussion *infra* Section IV.E.

9. Frederick Schauer, *Transparency in Three Dimensions*, 2011 U. ILL. L. REV. 1339, 1341 (2011).

10. Susanna Kim Ripken, *The Dangers and Drawbacks of the Disclosure Antidote: Toward a More Substantive Approach to Securities Regulation*, 58 BAYLOR L. REV. 139 (2006); Troy A. Paredes, *Blinded by the Light: Information*

offenders,¹¹ tire safety,¹² toxic pollution,¹³ and workplace chemical exposure,¹⁴ among many other species of conduct. Disclosure has even become a preferred method of regulation internationally.¹⁵

Although policymakers have relied on disclosure-based regulation for decades,¹⁶ it has evolved from peculiarity to regularity as the cost of disclosure online decreases and as public demand increases. But disclosure also has evolved in other important ways. For example, sometimes the real goal of disclosure is to persuade rather than inform.¹⁷ After all, is the Surgeon General's Warning on tobacco products meant to tell consumers something they do not already know? Or is it a suggestion not to smoke? Today, disclosure-based regulation is much less static (aimed narrowly at helping potential users of the information make better decisions) and more dynamic (aimed more broadly at trying to influence the disclosers' underlying behavior).¹⁸ Frequently, the real party being targeted by mandatory disclosure is not the consumer, but the discloser,¹⁹ under the Brandeisian logic that shining a light on undesired behavior will deter it, or at least make it more costly.

Overload and Its Consequences for Securities Regulation, 81 WASH. U. L. Q. 417 (2003).

11. Sex Offender Registration and Notification Act, 18 U.S.C. § 2250 (2006).

12. See Transportation Recall Enhancement, Accountability, and Documentation (TREAD) Act, § 12, 49 U.S.C. § 30118 (2000) (codified at 49 U.S.C. § 30170) (requiring manufacturers to disclose rollover risks for new vehicles). In 2007, Congress required that the information be placed on new car stickers. Safe, Accountable, Flexible, Efficient Transportation Equity Act: A Legacy for Users, Pub. L. No. 109-59, 119 Stat. 1144 (codified in scattered sections of 18, 23, and 49 U.S.C.).

13. Emergency Planning and Community Right-to-Know Act of 1986, 42 U.S.C. § 11001 (1986).

14. See OSHA Hazard Communication Standard, 29 C.F.R. § 1910.1200 (2009) (requiring material safety data sheets in the workplace).

15. ARCHONG FUNG ET AL., FULL DISCLOSURE: THE PERILS AND PROMISE OF TRANSPARENCY 127–50 (2007); ANNE-MARIE SLAUGHTER, A NEW WORLD ORDER 24–25 (2004) (comparing “regulation by information” efforts by the United States to those by the European Union and United Nations).

16. See, e.g., Howard Beales, Richard Craswell & Steven Salop, *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 491 (1981).

17. Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 744 (2011).

18. Richard Craswell, *Static Versus Dynamic Disclosures, and How Not to Judge Their Success or Failure*, 88 WASH. L. REV. 333, 339 (2013).

19. Ben-Shahar & Schneider, *supra* note 17.

Disclosure, then, remains in bloom. And this bloom is reflected in a relatively new species of disclosure—the searchable online database. In Part I below, I detail how agency databases derive from decades of federal policies promoting government transparency, particularly recent policies pushing the government to publish more information online. Online publication reached a crescendo with the Obama administration, which published an Open Government Memorandum on the President’s first day in office²⁰ and then promoted scores of other transparency projects, including the websites FOIA.gov and Data.gov. It is possible, if not likely, that data transparency by federal agencies will experience a diminuendo under the Trump administration,²¹ if not a more aggressive, weaponized use of disclosure aimed at particular parties.²²

Nevertheless, decades of groundwork has enabled not only a swell of government data initiatives, but also innovative nongovernmental uses of these data. Thousands of government data sets beget perhaps thousands more third-party websites, mobile applications, and other informational products that rely on government data. Perhaps the best example is the federal government’s decision in the 1980s to publish Global Positioning System (GPS) data for civilian use, which made possible the recent wave of applications that incorporate geospatial location data,²³ such as navigation, restaurant, and social media apps. Indeed, when President Obama announced his “Open Data Policy,” he remarked that “[t]his kind of

20. Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government, 74 Fed. Reg. 4685 (Jan. 21, 2009).

21. For possibilities of reversals in government data policy, see Clare Malone, *How Trump’s White House Could Mess with Government Data*, FIVETHIRTYEIGHT (Dec. 15, 2016, 6:29 AM), <http://fivethirtyeight.com/features/how-trumps-white-house-could-mess-with-government-data/> [<http://perma.cc/TQ6G-2G3Z>].

22. For example, when Donald Trump took to Twitter to criticize the contract price for Boeing to build a new Air Force One, Boeing stocks immediately dropped. See Phillip Bump, *Did Donald Trump Tank Boeing’s Stock Because He Was Mad about a News Article?*, WASH. POST (Dec. 6, 2016), <https://www.washingtonpost.com/news/the-fix/wp/2016/12/06/did-donald-trump-tank-boeings-stock-because-he-was-mad-about-a-news-article/> [<http://perma.cc/L9QM-6KQN>].

23. SCOTT PACE ET AL., *THE GLOBAL POSITIONING SYSTEM: ASSESSING NATIONAL POLICIES*, app. B GPS HISTORY, CHRONOLOGY, AND BUDGETS (Rand Corp. 1995), https://www.cs.cmu.edu/~sensing-sensors/readings/GPS_History-MR614.appb.pdf [<http://perma.cc/E7WZ-A267>]; *About*, HEALTHDATA.GOV, <http://www.healthdata.gov/content/about> (last visited July 16, 2017) [<http://perma.cc/B2L7-LZWV>].

innovation and ingenuity has the potential to transform the way we do almost everything.”²⁴ Lest readers discount this as hyperbole, disclosure policies are often burdened by their ambitions—being justified as promoting “autonomy, dignity, civility, community, citizenship, economic growth, and a variety of other virtues.”²⁵ Part II examines these aspirations in light of the emerging, somewhat sobering evidence.

Given the policy justifications, then, many rightly wonder who could possibly oppose providing more information to the public.²⁶ But as ubiquitous as disclosure has become, criticisms have emerged from scholars who doubt that it is “an unalloyed good.”²⁷ I consider the burdens of “regulation by database” in detail in Part III, including problems with fairness, accuracy, and efficacy. But for introductory purposes, Schauer captures the skepticism well:

Secrecy, privacy, anonymity, and confidentiality also have their virtues, and we can all understand why transparency is a far more desirable attribute for sunroom windows than it is for bathroom doors. At times, it seems that transparency is a prime example of the old adage that where you stand depends on where you sit.²⁸

But policymakers rarely question whether the burdens of disclosure outweigh its purported benefits.²⁹ One goal of this

24. Barack Obama, U.S. President, Remarks by the President at Applied Materials, Inc. - Austin, TX (May 9, 2013), <https://obamawhitehouse.archives.gov/photos-and-video/video/2013/05/09/president-obama-speaks-innovation-and-manufacturing> [<https://perma.cc/S46L-ZQ63>].

25. Ben-Shahar & Schneider, *supra* note 17, at 734.

26. FUNG ET AL., *supra* note 15, at xiii. Similarly, Fenster observes that “transparency appears to provide such a remarkable array of benefits that no right-thinking politician, administrator, policy wonk, or academic could be against it.” Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 888–89 (2006).

27. See, e.g., Schauer, *supra* note 9, at 1342; Ben-Shahar & Schneider, *supra* note 17; Amitai Etzioni, *Is Transparency the Best Disinfectant?*, 18 J. POL. PHIL. 389 (2010); Florencia Marotta-Wurgler, *Does Contract Disclosure Matter?*, 168 J. INST. & THEORETICAL ECON. 94 (2012); Karen E.C. Levy & David Merritt Johns, *When Open Data Is a Trojan Horse: The Weaponization of Transparency in Science and Governance*, BIG DATA & SOC’Y 1 (Jan.–June 2016), <http://journals.sagepub.com/doi/pdf/10.1177/2053951715621568> [<http://perma.cc/M4HP-VV4C>].

28. Schauer, *supra* note 9, at 1342.

29. Ben-Shahar & Schneider, *supra* note 17, at 683 (“[L]awmakers rarely inquire into the effectiveness or burden of disclosure.”).

Article is to help correct this asymmetry with regard to an increasingly important species of disclosure—the searchable federal database. In Part IV, I evaluate prominent databases published by six different agencies:

Agency Databases Evaluated

Agency	Database	Data
CFPB	Consumer Complaint Database	Complaints regarding financial product or service companies.
CPSC	SaferProducts.gov	“Reports of harm” for consumer products.
CMS	Hospital Compare, Physician Compare, etc.	Medicare quality of care metrics (such as mortality rates), with corresponding star ratings.
EPA	Toxics Release Inventory (TRI)	Production and release of roughly 650 dangerous chemicals by facility name, address.
FDA	FDA Adverse Event Reporting System (FAERS), Manufacturer and User Facility Device Experience (MAUDE)	Adverse events associated with drugs (FAERS) and medical devices (MAUDE) reported to the FDA by manufacturers, health providers, consumers.
FEC	Campaign Finance Disclosure Portal	Mandatory reports made by federal candidates, parties, committees, donors, lobbyists, others.

Part IV considers which of these databases succeed in producing reliable, usable data, and which have been able to influence the underlying behavior being tracked. I also consider how these databases suffer from different types of flaws, including incomplete or inaccurate data, unfriendly user interfaces, or databases whose costs of collection, maintenance, and presentation likely outweigh their utility.

Part V offers thoughts on how policymakers can design databases for more optimal disclosure, focusing on the inputs

and outputs of published data, including procedural safeguards to help ensure the quality and reliability of data. The recommendations include pre- and post-publication procedures for adjudicating “contested” data, evaluating administrative law problems that might arise when databases are implicated in enforcement actions, offering ideas on how to characterize and present the data fairly and accurately, and drawing lessons for agencies considering whether to publish “raw” or “polished” data, as well as “big” or “small” data.

Together, these recommendations envision a decidedly modern role for the government as a “data steward”³⁰ rather than merely as a data source or publisher. For example, there are smart ways that the government can help gather and generate more data—making data “bigger.”³¹ But for some information, it might be preferable for the government to distill the data and make it more user-friendly—making data “smaller.”³² Either way, federal agencies can help ensure that data collection practices are fair, that data that purport to be accurate and objective meet those standards, and that the sources and any important context or limitations for the data are communicated clearly to users. In particular, policymakers should resist the notion that more data is always better data. Quality matters more than quantity. Part V thus builds on my recent work for the Administrative Conference of the United States (ACUS), which recently adopted a narrower set of recommendations on consumer complaint databases.³³ We might refer to these recommendations collectively as “Good Government Data Practices.”

As the Article begins, consider the stakes. Successful data policies can help ensure healthy markets, empower consumers, inform citizens, and even influence the underlying conduct of disclosers. Failed data policies, conversely, can produce incomplete or “gerrymandered” information, create a false

30. Kristin Madison, *Health Care Decisions in the New Era of Health Care Reform: Health Regulators as Data Stewards*, 92 N.C. L. REV. 1605, 1607–08 (2014).

31. *Id.* at 1627–28.

32. *Id.*

33. ACUS Adoption of Recommendations, 81 Fed. Reg. 40,259 (June 21, 2016); ADMIN. CONFERENCE OF THE U.S., *Agency Information Dissemination in the Internet Era*, <https://www.acus.gov/research-projects/agency-information-dissemination-internet-era> (last visited July 16, 2017) [<http://perma.cc/HP2B-BWHY>] [hereinafter ACUS].

sense of security, waste resources, undermine public trust, and even put lives at risk.³⁴ To design databases for optimal disclosure, we must also appreciate that disclosure is an exercise of power, that “for one person or institution to have information about another is for the former to have power over the latter.”³⁵ Similarly, “transparency” can be recharacterized as “adverse transparency” if the information is unflattering or harmful to the subject in some way.³⁶ How, then, can policymakers exercise this power responsibly and fairly? If we value disclosure and accept it as a baseline, how do we best manage it? As federal data is used more and more to achieve regulatory ends, both the means and ends require more purposeful policies.

I. FROM OPEN GOVERNMENT TO OPEN INDUSTRY

Recent efforts to shine light on the activities of the regulated derive from very old efforts to shine light on the activities of regulators themselves. For decades, perhaps even centuries,³⁷ citizens have pushed the U.S. government to be more open and transparent. Indeed, the long arc toward government transparency is a defining hallmark of American administrative law,³⁸ sitting comfortably “among the pantheon of great political virtues.”³⁹ But today’s transparency emerged only after decades of reforms, arriving roughly in four waves. The 1930s brought efforts to publish so-called “secret laws” generated by agencies during the New Deal. The 1940s brought the Administrative Procedure Act’s mandates to give regulated parties advanced notice of agency actions. The 1960s and 1970s

34. FUNG ET AL., *supra* note 15, at 172.

35. Schauer, *supra* note 9, at 1347.

36. Sarah Taylor Roller et al., *FDA’s Expanding Postmarket Authority to Monitor and Publicize Food and Consumer Health Product Risks: The Need for Procedural Safeguards to Reduce “Transparency” Policy Harms in the Post-9/11 Regulatory Environment*, 64 FOOD & DRUG L.J. 577, 597 (2009).

37. Article I, section 5 of the Constitution requires each chamber of Congress to “keep a Journal of its Proceedings, and from time to time to publish the same,” which some view as a deliberate departure from the secrecy practiced by the British Parliament. U.S. CONST. art. I § 5; James J. Brudney, *Canon Shortfalls and the Virtues of Political Branch Interpretive Assets*, 98 CAL. L. REV. 1199, 1218 (2010).

38. William Funk, *Public Participation and Transparency in Administrative Law—Three Examples as an Object Lesson*, 61 ADMIN. L. REV. 171 (2009).

39. Fenster, *supra* note 26, at 888.

introduced FOIA and the era of transparency by request. And the 1990s and 2000s introduced mandates for agencies to post information on the internet, establishing important agency norms of online publication.

A. *Publishing “Secret Laws”*

The New Deal birthed a generation of new executive agencies and corresponding regulations. But agencies published their regulations at will, if at all.⁴⁰ One of the earliest efforts to address the lack of transparency among federal agencies was the Federal Register Act of 1935 (the 1935 Act),⁴¹ which created the now-familiar daily gazette of executive documents. Before the 1935 Act, executive branch agencies “would each publish their own regulations in various separate publications, be they gazettes, bulletins, rulings, digests, pamphlets, notices, codes, certificates, orders, and the like.”⁴²

The Act was motivated in part by the famous “hot oil” case, *Panama Refining*.⁴³ The “hot oil” law was part of the National Industrial Recovery Act, the flagship New Deal bill passed during the Great Depression.⁴⁴ The National Industrial Recovery Act authorized President Roosevelt, via the Secretary of the Interior, to limit oil production and stabilize prices during the discovery of vast new oil fields in Texas.⁴⁵ During litigation over the new authority, “the government was embarrassed to admit that a reexamination of the relevant documents (which were not publicly available) had revealed that the Secretary had inadvertently revoked the relevant regulation before the lawsuit had been filed.”⁴⁶ Just weeks before oral argument in the case, Erwin Griswold published a

40. Rick McKinney, *A Research Guide to the Federal Register and the Code of Federal Regulations*, 46 LAW LIBR. INSIGHTS 10, 10 (2002), <http://www.llsdc.org/assets/sourcebook/fall02.pdf> [<http://perma.cc/5JG4-X54V>].

41. 41 U.S.C. § 1501 (1935).

42. McKinney, *supra* note 40, at 10.

43. *Panama Ref. Co. v. Ryan*, 293 U.S. 388 (1935).

44. National Industrial Recovery Act, Pub. L. No. 73-67, 48 Stat. 195 (1933).

45. For a brief history on the discovery of new fields in Texas during that time, see ROBERT A. CARO, *THE PATH TO POWER: THE YEARS OF LYNDON JOHNSON* 612–15 (1981).

46. PETER STRAUSS ET AL., *GELLHORN AND BYSE’S ADMINISTRATIVE LAW: CASES AND COMMENTS* 446 (11th ed. 2011).

law review article, *Government in Ignorance of the Law*, arguing for a Federal-Register-like system to publish executive branch laws.⁴⁷ The “furor” over the case⁴⁸ reflected longstanding and “widespread dissatisfaction with the unsystematic manner in which executive orders, agency regulations, and similar materials were being made available to the public.”⁴⁹ As the federal government swelled with new agencies and new regulations, frustration reached even high-level government officials, who found it difficult, if not impossible, to locate what became known as “secret laws.”⁵⁰ Before 1935, agencies did not even have to publish the regulations they imposed.⁵¹

B. *The APA and Notice*

Publication requirements were further embedded in American law by the Administrative Procedure Act of 1946 (APA).⁵² The APA was a response, in part, to complaints from industry that administrative agencies during the New Deal

47. Erwin N. Griswold, *Government in Ignorance of the Law—A Plea for Better Publication of Executive Legislation*, 48 HARV. L. REV. 198 (1934). Griswold, the future U.S. Solicitor General and Dean of Harvard Law School, most likely anticipated the *Panama Refining* decision. He was an attorney at the Solicitor General’s office until 1934 during the briefing of the case. His article was published the same month (December 1934) as oral argument in the case. STRAUSS ET AL., *supra* note 46, at 446 n.4.

48. STRAUSS ET AL., *supra* note 46, at 446 n.4 (citing Mary Whisner, *A Manual “to Inform Every Citizen,”* 99 LAW LIBR. J. 159, 160 (2007); MORRIS L. COHEN, ROBERT C. BERRING, & KENT C. OLSON, *HOW TO FIND THE LAW* 265 (9th ed. 1989)).

49. *See Cervase v. Office of the Fed. Register*, 580 F.2d 1166, 1169 (3d Cir. 1978) (“The basic object of this statutory reform was to eliminate secret law.”).

50. OFFICE OF THE FED. REGISTER, *THE FEDERAL REGISTER MARCH 14, 1936 – MARCH 14, 2006*, at 2 (2006), https://www.federalregister.gov/uploads/2011/01/fr_history.pdf [<http://perma.cc/A3L5-B6HS>]; STRAUSS ET AL., *supra* note 46, at 445–46.

51. Funk, *supra* note 38, at 172. Decades later, in 1993, Congress required the Government Printing Office (GPO) to make the Federal Register available on the Internet. Government Printing Office Information Access Enhancement Act of 1993, Pub. L. No. 103-40, 107 Stat. 112 (codified as amended at 44 U.S.C. §§ 4101–4104).

52. Administrative Procedure Act (APA), Pub. L. No. 79-404, 60 Stat. 237 (1946) (codified as amended in scattered sections of 5 U.S.C.). States later passed their own administrative procedure acts, often modeled on the federal APA, thus incorporating the same publication requirements and principles. *See* MICHAEL A. ASIMOW & RONALD M. LEVIN, *STATE AND FEDERAL ADMINISTRATIVE LAW* 3–5 (4th ed. 2014).

were too opaque and insular, particularly towards the private interests most affected by regulation.⁵³ The original APA, in section 3, required agencies to publish important materials in the Federal Register, and in fact prohibited agencies from enforcing rules not published there.⁵⁴ New APA procedures creating “notice and comment” rulemaking have since become a hallmark of citizen participation in government.⁵⁵ Today, the APA requires agencies to publish a wide variety of information in the Federal Register, including basic information about their organizational structure, procedures, and substantive rules.⁵⁶

However, many eventually came to view APA section 3 more as a tool to *withhold* information than disclose it.⁵⁷ And even between the Federal Register Act and the APA, a significant portion of agency documents—guidance, opinions, and other important “soft law” adopted by agencies—were not available in the Federal Register. Just as before 1935, such documents were accessible only on a haphazard basis.⁵⁸

Moreover, the APA’s disclosure provisions were largely designed to give notice to those whose legal rights were directly affected by the agency action, which were “almost invariably businesses.”⁵⁹ Indeed, notice-and-comment rulemaking procedures were premised on the view that regulated parties, rather than the public at large, should be given notice and an opportunity to comment on proposed rules.⁶⁰ As Bill Funk observes, the APA addressed participation in rulemaking to “interested persons,” meaning those with a “direct and palpable interest,”⁶¹ and required that public records be made available to persons “properly and directly concerned,” rather than the

53. Funk, *supra* note 38, at 172–73, 178.

54. APA § 3, 60 Stat. at 238.

55. 5 U.S.C. § 553 (2012); Jennifer Shkabatur, *Transparency With(out) Accountability: Open Government in the United States*, 31 YALE L. & POL’Y REV. 79, 85 (2012); Beth Simone Noveck, *The Electronic Revolution in Rulemaking*, 53 EMORY L.J. 433, 517 (2004).

56. 5 U.S.C. § 552(a)(1) (2012).

57. STRAUSS ET AL., *supra* note 46, at 451.

58. *Id.*

59. Funk, *supra* note 38, at 173.

60. *Id.*; ATTORNEY GEN.’S COMM. ON ADMIN. PROCEDURE, DEP’T OF JUSTICE, FINAL REPORT OF THE ATTORNEY GENERAL’S COMMITTEE ON ADMINISTRATIVE PROCEDURE 2 (1941).

61. Funk, *supra* note 38, at 174.

general public.⁶² Thus, the APA's disclosure provisions were aimed to inform regulated parties, not shine a light on them, as became the focus decades later.

C. *FOIA and Transparency by Request*

If the APA is viewed as a reaction to agency hostility towards regulated businesses, then the 1960s and 1970s could be viewed as a movement toward recognizing the public interest on equal footing.⁶³ The modern open government movement really began in 1967, when Congress passed the Freedom of Information Act (FOIA),⁶⁴ requiring agencies to index and make public vast amounts of materials not published in the Federal Register. Like the APA, FOIA was motivated in part by the desire to ensure an informed citizenry.⁶⁵ President Lyndon Johnson, who signed the bill reluctantly and only under pressure from the press corps,⁶⁶ noted that FOIA “springs from one of our most essential principles: a democracy works best when the people have all the information that the security of the nation will permit.”⁶⁷ Half a century later, modern scholars still acknowledge FOIA's importance to our democratic government.⁶⁸ In requiring the government to make its records available upon request unless specifically exempted,

62. APA, Pub. L. No. 79-404, § 3(c), 60 Stat. 237, 238 (1946) (codified as amended in scattered sections of 5 U.S.C.). Funk explains that these goals derive in part from the canonical work by James Landis, *The Administrative Process*. Funk, *supra* note 38, at 177 (citing JAMES M. LANDIS, *THE ADMINISTRATIVE PROCESS* (1938)).

63. Funk, *supra* note 38, at 178–80.

64. Freedom of Information Act (FOIA), Pub. L. No 90-23, 81 Stat. 54 (1967) (current version at 5 U.S.C. § 552 (1996)). The original FOIA was enacted in 1966, Pub. L. No. 89-487, 80 Stat. 250, but was repealed and replaced after Congress enacted Title 5 of the U.S. Code into positive law by the 1967 version, which was identical in substance. Vladeck has characterized FOIA as “truly an experiment in open government.” David C. Vladeck, *Information Access – Surveying the Current Legal Landscape of Federal Right-to-Know Laws*, 86 TEX. L. REV. 1787, 1795 (2008).

65. *See, e.g.*, 112 CONG. REC. H13007 (daily ed. June 20, 1966) (statement of Rep. Moss).

66. Vladeck, *supra* note 64, at 1798.

67. President Lyndon B. Johnson, Presidential Statement on Signing the Freedom of Information Act (July 4, 1966), <http://nsarchive.gwu.edu/NSAEBB/NSAEBB194/Document%2031.pdf> [<https://perma.cc/R6YQ-E7XC>].

68. Vladeck, *supra* note 64 (noting that FOIA “embodies the ideal that information is the lifeblood of democracy”).

FOIA created a “strong presumption in favor of disclosure.”⁶⁹ In fact, it reversed the burden in the original APA that opened access to government records only if the requester could demonstrate a compelling need.⁷⁰

FOIA, of course, has been criticized for falling short of its lofty goals of pursuing democracy through transparency and accountability.⁷¹ A major complaint is that FOIA produces transparency only by request.⁷² It imposes few affirmative disclosure obligations on agencies, and relies on a complex framework that often requires relatively sophisticated private intermediaries with sufficient “time, money, and expertise” to “press a recalcitrant administration for disclosure.”⁷³ Scholars have also criticized FOIA for being too malleable, particularly under administrations that construe its disclosure requirements narrowly and exemptions broadly (the George W. Bush administration is frequently cited).⁷⁴ As such, FOIA envisions passive disclosure by agencies rather than active disclosure or publication of the information they hold.⁷⁵ Moreover, the basic premise of FOIA’s “request-and-wait-for-a-response approach” is seen as obsolete in the Internet era.⁷⁶

D. *The Internet Era*

In the 1990s, as the internet came of age, a series of laws pushed the government to use it. For example, the Paperwork

69. 5 U.S.C. § 552(a)–(b); U.S. Dep’t of State v. Ray, 502 U.S. 164, 173 (1991).

70. 4 U.S.C. § 1002(c) (1964), *amended by* 5 U.S.C. § 552(a)(1)–(4) (Supp. III 1964).

71. Shkabatur, *supra* note 55, at 88–89 (citing several articles and at least one law, the Openness Promotes Effectiveness in Our Nation (OPEN) Government Act of 2007, Pub. L. No. 110-175, 121 Stat. 2524 (codified in scattered sections of 5 U.S.C.)).

72. Vladeck, *supra* note 64, at 1789.

73. Seth F. Kreimer, *The Freedom of Information Act and the Ecology of Transparency*, 10 U. PA. J. CONST. L. 1011, 1020 (2008); Shkabatur, *supra* note 55, at 89; Vladeck, *supra* note 64, at 1789.

74. Scholars often point to the George W. Bush administration on this point. President Bush instructed federal agencies to deny FOIA requests when they could invoke a “sound legal basis.” Memorandum from John Ashcroft, U.S. Att’y Gen., to the Heads of All Fed. Dep’ts & Agencies (Oct. 12, 2001); Vladeck, *supra* note 64, at 1790; Shkabatur, *supra* note 55, at 89.

75. Nathan Cortez, *Adverse Publicity by Administrative Agencies in the Internet Era*, 2011 BYU L. REV. 1371, 1438–39 (2011).

76. Vladeck, *supra* note 64, at 1792–93.

Reduction Act of 1995,⁷⁷ the Electronic Freedom of Information Act of 1996 (sometimes called “E-FOIA”),⁷⁸ and the Government Paperwork Elimination Act of 1998⁷⁹ directed federal agencies to use the internet to publish more information online and to “improve the productivity, efficiency, and effectiveness of Federal programs.”⁸⁰ In particular, E-FOIA required agencies to publish online their final opinions and orders, as well as records likely to be requested,⁸¹ thus spawning the surge in agency online “reading rooms.”⁸²

During this time, regulators began to rely more on mandatory reporting to inform agency actions, including both rulemaking and enforcement. Somewhat quickly, internet technologies reduced the costs of gathering and disseminating such information online, which increased both public demand and expectations for agency records. Thus, federal agencies began to publish copious amounts of information not just about their own activities, but about regulated parties as well.⁸³

In the 2000s, as federal agencies built sprawling websites,⁸⁴ a new generation of laws pushed for even more online disclosure. For example, the E-Government Act of 2002 required federal agencies to post more information online and make it more accessible through improved organization.⁸⁵ The Act also created the Office of Electronic Government within the Office of Management and Budget (OMB), headed by a Chief Information Officer that would coordinate with agencies through a Council.⁸⁶ As with prior laws, the stated goal of the E-Government Act was to encourage the federal government to enhance public access to information and government services,

77. Pub. L. No. 104-13, § 2, 109 Stat. 163, 167 (codified at 44 U.S.C. § 3504(h)(5)).

78. Pub. L. No. 101-231, 110 Stat. 347 (codified at 5 U.S.C. § 552).

79. Pub. L. No. 105-277, div. C, tit. XVII, § 1702, 112 Stat. 2681 (codified at 44 U.S.C. § 3504).

80. Pub. L. No. 104-13, § 2, 109 Stat. 163, 167 (codified at 44 U.S.C. § 3504(h)(5)).

81. 5 U.S.C. § 552(a)(2)(D).

82. Presidential Statement on Signing the Electronic Freedom of Information Act Amendments of 1996, 32 WEEKLY COMP. PRES. DOC. 1949 (Oct. 2, 1996).

83. James O'Reilly, *Libels on Government Websites: Exploring Remedies for Federal Internet Defamation*, 55 ADMIN. L. REV. 507 (2003).

84. *Id.*; Cortez, *supra* note 75, at 1393.

85. E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899 (codified as amended in scattered sections of the U.S. Code).

86. *Id.*

this time using internet technologies.⁸⁷ In 2007, the Open Government Act addressed various frustrations with FOIA, requiring new agency procedures and new public liaison offices designed to address persistent agency delays in responding to FOIA requests.⁸⁸

Likewise, during the 2000s, Congress also passed laws requiring more transparency in federal spending, directing the OMB to publish online the details of federal grants, loans, and contracts.⁸⁹ Today, the public can search the federal website USAspending.gov to view entities that have received federal money,⁹⁰ or more narrowly focused sites like Recovery.gov to see how the federal government has spent money from the economic stimulus package of 2009.⁹¹ The former includes a searchable database with the name and location of the entity receiving federal money, the amount received, the type of transaction, the funding body, the purpose of the funding, and other information.⁹² As Vladeck notes, the site “was able to piggyback on the work of OMB Watch, a nonprofit watchdog organization that with foundation support had already constructed a comprehensive, searchable database that is also

87. *Id.* § 101(a), 116 Stat. at 2902 (defining “electronic government”).

88. Open Government Act of 2007, Pub. L. No. 110-175, 121 Stat. 2524 (codified in scattered sections of 5 U.S.C. § 552).

89. E-Government Act of 2002; Federal Funding Accountability and Transparency Act of 2006, Pub. L. No. 109-282, 120 Stat. 1186 (codified at 31 U.S.C. § 6101).

90. *See* USASPENDING.GOV, <http://www.usaspending.gov> (last visited Aug. 3, 2017) [<https://perma.cc/2LDE-WR48>].

91. U.S. RECOVERY ACCOUNTABILITY AND TRANS. BD., <http://www.recovery.gov>. The site and its data were removed after the Recovery Accountability and Transparency Board declined to renew a license with Dun & Bradstreet, a firm that assigned identification numbers for all entities doing business with the U.S. Government. *See* Christian Davenport, *Data on \$800 Billion in Stimulus Spending Will Disappear This Year. Here is Why.*, WASH. POST (Sept. 9, 2014), https://www.washingtonpost.com/business/economy/data-on-800-billion-in-stimulus-spending-will-disappear-this-year-here-is-why/2014/09/09/ad277ff4-350a-11e4-8f02-03c644b2d7d0_story.html [<https://perma.cc/V62M-ZMM5>]. In 2012, the U.S. House passed a bill that would publish even more online data about federal spending, and would have created a Federal Accountability and Spending Transparency Commission to implement the new provisions. *See* Digital Accountability and Transparency (DATA) Act, H.R. 2146, 112th Cong. (2012). The bill died in the Senate. *See* THE LIBRARY OF CONGRESS, *Bill Summary & Status, 112th Congress (2011-12), H.R. 2146*, <http://thomas.loc.gov/cgi-bin/bdquery/z?d112:h.r.02146> (last visited July 15, 2017) [<https://perma.cc/N256-82W7>].

92. Pub. L. No. 109-282, § 2(b)(1), 120 Stat. 1187; *see also* USASPENDING.GOV, *supra* note 90.

available free of charge to the public.”⁹³ This work reflects a marked shift from focusing on information about the government to information about private parties.

In 2001, Congress passed an important but less frequently discussed law, the Information Quality Act (IQA), sometimes referred to as the Data Quality Act (DQA).⁹⁴ The Act required the OMB to issue government-wide guidelines for “ensuring and maximizing the quality, objectivity, utility, and integrity of information . . . disseminated by the government.”⁹⁵ It also required the OMB to “establish administrative mechanisms allowing affected persons to seek and obtain correction of information maintained and disseminated by the agency” that does not meet those standards.⁹⁶ In 2002, the OMB finalized guidelines implementing the Act,⁹⁷ followed by agencies issuing their own guidelines.⁹⁸

These IQA guidelines would seem well suited to regulate the quality of information posted in agency databases. Yet the Act’s application to databases is highly unclear, and probably varies by database. The broad wording of the IQA states that the OMB guidelines should apply to agency “dissemination of public information, regardless of the form or format.”⁹⁹ And the OMB guidelines define “information” as “any communication or representation of knowledge such as facts or data, in any

93. Vladeck, *supra* note 64, at 1829–30 (citing *About*, FEDSPENDING.ORG, <http://www.fedspending.org/aboutthissite.php> (last visited August 1, 2017) [<https://perma.cc/QPC2-7ZT5>]). Note that OMB Watch later changed its name to the Center for Effective Government, but ended operations in March 2016, transferring most of its resources to the Project on Government Oversight (POGO). See CTR. FOR EFFECTIVE GOV’T, <http://www.foreffectivegov.org> (last visited June 6, 2017) [<https://perma.cc/7GF7-LGSD>].

94. Treasury and General Government Appropriations Act for Fiscal Year 2001, Pub. L. No. 106-554, § 515, 114 Stat. 2763, 2763A-153-54 (2001); 44 U.S.C. § 3516 (2012).

95. Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515. The Information Quality Act built on earlier requirements in the Paperwork Reduction Act that addressed information dissemination. See Paperwork Reduction Act of 1995, Pub. L. No. 104-13, 109 Stat. 163, 168.

96. Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515.

97. 66 Fed. Reg. 34,489 (June 28, 2001) (proposed guidelines); 66 Fed. Reg. 49,718 (Sept. 28, 2001); 67 Fed. Reg. 369 (Jan. 3, 2002); 67 Fed. Reg. 8452 (Feb. 22, 2002).

98. OFFICE OF MGMT. & BUDGET, *Agency Information Quality Guidelines*, https://obamawhitehouse.archives.gov/omb/inforeg_agency_info_quality_links/ (last visited July 15, 2015) [<https://perma.cc/8YZZ-PKRF>].

99. 44 U.S.C. § 3504(d)(1).

medium or form,”¹⁰⁰ including “information that an agency disseminates from a web page.”¹⁰¹ However, the OMB excludes from coverage “opinions, where the agency’s presentation makes it clear that what is being offered is someone’s opinion rather than fact or the agency’s views.”¹⁰² The guidelines also exempt “adjudicative processes.”¹⁰³ These exemptions might thus exclude important agency databases, such as the CFPB’s Consumer Complaint Database.¹⁰⁴

All this is prelude, however, to the Obama administration and its efforts toward open government. On his first full day in office, President Obama published the Open Government Memorandum,¹⁰⁵ as well as a memorandum on FOIA.¹⁰⁶ Although various internet-driven transparency initiatives emerged during the Clinton and Bush administrations, the two Obama documents were viewed as a gesture toward openness and a turn from the secrecy that characterized the Bush administration.¹⁰⁷

In December 2009, the OMB published the Open Government Directive,¹⁰⁸ following on President Obama’s Open Government Memorandum, urging agencies to “take prompt steps to expand access to information by making it available online in open formats.”¹⁰⁹ The directive required each agency, within 45 days, to “identify and publish online in an open format at least three high-value data sets . . . on Data.gov” that

100. 67 Fed. Reg. 8452, 8458–59.

101. *Id.* at 8460.

102. 67 Fed. Reg. 369, 377.

103. *Id.*

104. See discussion *infra* Section IV.C.

105. Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government (Jan. 21, 2009), 74 Fed. Reg. 4685 (Jan. 21, 2009). As Shkabatur notes, “[d]ozens of other countries” have followed this example. Shkabatur, *supra* note 55, at 80 (citing OPEN GOVERNMENT PARTNERSHIP, <http://www.opengovpartnership.org> (last visited July 15, 2017) [<https://perma.cc/R85P-AWL5>]).

106. Memorandum for the Heads of Executive Departments and Agencies: Freedom of Information Act, 74 Fed. Reg. 4683 (Jan. 21, 2009).

107. STRAUSS ET AL., *supra* note 46, at 441. Of course, even the Obama administration has been criticized for rejecting more transparency in matters of terrorism and national security. Jeff Kahn, *Terrorist Watchlists*, in CAMBRIDGE HANDBOOK OF SURVEILLANCE LAW (David Gray & Stephen E. Henderson eds., Cambridge Univ. Press, 2017).

108. Memorandum from Peter R. Orszag, Dir., Office of Mgmt. & Budget, to the Heads of Exec. Dep’ts and Agencies (Dec. 8, 2009) (on file with author).

109. *Id.* at 2.

had previously not been available.¹¹⁰ Within sixty days, each agency was to create an Open Government web page. Today, eighty different federal agencies and subagencies have posted 196,284 datasets on Data.gov.¹¹¹

In 2011, the Justice Department created FOIA.gov, a website that publicizes data on how agencies have performed their FOIA duties.¹¹² The searchable online database displays the number of FOIA requests received by each agency, the disposition of those requests, and the current backlog. Ironically, like other mandatory disclosure regimes, the Justice Department is using “naming and shaming” to encourage agencies to be more responsive to FOIA requests.¹¹³ Still, scholars question how effectively “naming and shaming” is at convincing under-performing agencies to increase their responsiveness to FOIA requests.¹¹⁴

As this history shows, federal transparency efforts gradually evolved from general right-to-know laws like FOIA, aimed at increasing transparency in the government itself, to disclosure of information held by the government regarding the activities of corporations and regulated entities.¹¹⁵ Thus, the current gestalt that pursues transparency from industry owes much to earlier right-to-know efforts that pursued transparency from government.¹¹⁶ Moreover, as with so many other things, information technology has enabled the use of databases and disclosure as a regulatory tool. Internet technologies are being used to mine the data of countless industries and activities, post them in the public domain, and make them accessible through searching, sorting, and other data-sifting tools.

110. *Id.*

111. *Federal Agency Participation*, *supra* note 3.

112. U.S. DEPT OF JUSTICE, *What is FOIA?*, FOIA.GOV, <http://www.foia.gov> (last visited Jan. 12, 2017) [<https://perma.cc/AQ6A-PE9Q>].

113. Shkabatur, *supra* note 55, at 100.

114. *Id.*

115. *See also* FUNG ET AL., *supra* note 15, at xii–xiii. They call this “targeted transparency.” *Id.*

116. *Id.* at 28. Of course, one could support more transparency from government without supporting more transparency from industry, and vice versa. But the two trends seem to be part of the same historical arc towards more public reporting and openness by regulatory agencies.

II. DATA AND ITS ASPIRATIONS

Agency databases have become ubiquitous in part because they appeal to so many of our intuitions about how government, markets, and regulation should work. So much so, in fact, that disclosure skeptics criticize the optimists for too often posing it as a panacea—a Swiss Army policy “intended to promote autonomy, dignity, civility, community, citizenship, economic growth, and a variety of other virtues.”¹¹⁷ Contemporary scholarship, of course, has focused on the many ways in which internet technologies have facilitated communication between the government and the public.¹¹⁸ But after years of scholarly praise of the internet’s role in facilitating transparency, accountability, and democracy, inevitable critiques have emerged.¹¹⁹ Still, disclosure frequently is offered as a tool that can achieve market, regulatory, and democratic ideals.¹²⁰

A. *Market Ideals*

A frequently invoked rationale for regulatory disclosures is that disclosure can improve consumer decision-making, facilitate markets, and “protect the naïve from the sophisticated.”¹²¹ Various disclosure regimes, at their heart, try to resolve the famous “lemons problem” framed by George Akerloff, who argued that in markets with information asymmetries between buyers and sellers, sellers may have an incentive to sell inferior products or services, which can

117. Ben-Shahar & Schneider, *supra* note 17, at 734.

118. For just a small sample, see, e.g., BRUCE BIMBER, *INFORMATION AND AMERICAN DEMOCRACY: TECHNOLOGY AND THE EVOLUTION OF POLITICAL POWER* (W. Lance Bennett & Robert M. Entman eds., 2003); STEPHEN COLEMAN & JAY G. BLUMLER, *THE INTERNET AND DEMOCRATIC CITIZENSHIP: THEORY, PRACTICE AND POLICY* (2009).

119. See, e.g., MATTHEW HINDMAN, *THE MYTH OF DIGITAL DEMOCRACY* (2003); EVGENY MOROZOV, *THE NET DELUSION: THE DARK SIDE OF INTERNET FREEDOM* (2011).

120. Indeed, Sage evaluated these types of justifications almost twenty years ago, characterizing them somewhat differently as four separate rationales (competition, agency, performance, and democratic). See Sage, *supra* note 5, at 1710–11. He also observes, however, that disclosure rationales can often be contradictory or at least in tension. *Id.*

121. Ben-Shahar & Schneider, *supra* note 17, at 649.

undermine the market.¹²² By requiring disclosure, the government can correct these asymmetries and facilitate efficient markets. Schauer calls this “transparency as efficiency”—the idea that freely available information “is precisely what makes markets operate effectively.”¹²³

At their best, agency databases can inspire a “race to the top” by encouraging firms to compete based on their published activities. One of the original aspirations for the CFPB’s Consumer Complaint Database was to encourage companies to use the data to publicize how well they respond to consumer complaints compared to competitors.¹²⁴ The Bureau points to this phenomenon in the airline industry, where airlines use data by the Department of Transportation and FAA to market their low rates of passenger complaints compared to competitors, and where third party airline ratings systems make use of the same government data.¹²⁵ The Bureau concludes that after the data is made public, “The marketplace of ideas then does the rest.”¹²⁶

Disclosure thus satisfies both our free-market intuitions¹²⁷ and the policymaker’s urge to do *something*. In the law that created the CFPB, Congress repeatedly asserted that the Bureau would publish information that helped consumers make more informed choices about financial products and services¹²⁸—a refrain repeated frequently by the Bureau in its

122. George A. Akerloff, *The Market for “Lemons”: Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488 (1970) (discussing the used car market as an example).

123. Schauer, *supra* note 9, at 1350.

124. Disclosure of Certain Credit Card Complaint Data, 76 Fed. Reg. 76,628, 76,630 n.9 (Dec. 8, 2011).

125. *Id.* at 76,631.

126. *Id.* Of course, airlines now face a new form of naming and shaming from customers themselves, who can record and publish examples of poor customer service. See, e.g., Avi Selk, *A Man Wouldn’t Leave an Overbooked United Flight. So He Was Dragged Off, Battered and Limp.*, WASH. POST (Apr. 10, 2017), <https://www.washingtonpost.com/news/dr-gridlock/wp/2017/04/10/a-man-wouldnt-leave-an-overbooked-united-flight-so-he-was-dragged-off-battered-and-limp/> [<https://perma.cc/8V5B-VEE6>].

127. Ben-Shahar & Schneider, *supra* note 17, at 681.

128. Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, § 1021, 124 Stat. 1979 (2010) (codified as 12 U.S.C. §§ 5511(b)(1), (5)) (tasking the Bureau with providing consumers “timely and understandable information to make responsible decisions about financial transactions” and help the market “operate transparently and efficiently”).

own publications.¹²⁹ Consumer advocates, in fact, have encouraged the Bureau's database efforts by arguing that "disclosure is one of the best tools government agencies can use."¹³⁰

A related consumer-centered ideal served by disclosure is autonomy. As Ben-Shahar and Schneider argue, mandated disclosure is alluring because "[i]t supposes that people make better decisions for themselves than anyone can make for them and that people are entitled to freedom in making decisions."¹³¹ Countless mandatory disclosure laws rely on this logic.¹³² The CFPB's complaint data, the CPSC's product safety data, and many other data sets are predicated on consumers using the data to vote with their wallets, avoiding substandard performers.¹³³

Finally, corporations and industry groups often "urge greater transparency as an alternative to allegedly more heavy-handed regulation."¹³⁴ Of course, many scholars embrace this view as well.¹³⁵ As Archon Fung and colleagues emphasize, the "ingeniousness" of regulation via disclosure "lies in its mobilization of individual choice, market forces, and participatory democracy through relatively light-handed government action."¹³⁶ Regulation by disclosure thus appeals across both political and ideological spectra.

B. Regulatory Ideals

A second aspiration of disclosure is to achieve regulatory ends—using publication to preempt or at least deter undesired behavior. Corporate and securities law, for example, rely heavily on disclosure of company holdings and transactions, with the idea that corporations whose dealings are transparent

129. See, e.g., Disclosure of Consumer Complaint Data, 78 Fed. Reg. 21,218, 21,225 (Apr. 10, 2013).

130. *Id.* at 21,220.

131. Ben-Shahar & Schneider, *supra* note 17, at 681.

132. *Id.* (citing examples from the FTC and others).

133. See discussion *infra* Sections IV.C and IV.D.

134. Schauer, *supra* note 9, at 1341 (citing L. Gordon Crovitz, Opinion, *Transparency is More Powerful Than Regulation*, WALL ST. J., Mar. 30, 2009, at A21; Richard H. Thaler & Cass R. Sunstein, Opinion, *Disclosure is the Best Kind of Credit Regulation*, WALL ST. J., Aug. 13, 2008, at A17).

135. See, e.g., Crovitz, *supra* note 134; Thaler & Sunstein, *supra* note 134.

136. FUNG ET AL., *supra* note 15, at 5.

and publicly accessible will think twice before acting in ways that harm investors.¹³⁷ Similarly, requiring hospitals to publish mortality rates is really a device to encourage hospitals to *reduce* mortality rates.¹³⁸ Again, the same logic motivates many disclosure regimes, with a long lineage. In 1796, Jeremy Bentham observed that “the more strictly we are watched, the better we behave.”¹³⁹

Agencies also frequently use databases to publish compliance and enforcement data.¹⁴⁰ Scholars have called for agencies to actively publish enforcement records that are available under FOIA but must be requested.¹⁴¹ For example, David Vladeck argues that Congress should require the OMB to compile enforcement records in a searchable database to “permit the public to track repeat-offender corporations in the same way the public can now track grants and contracts given to the same corporate recipients.”¹⁴² For years, a nonprofit based at Syracuse University, called the Transactional Records Access Clearinghouse (TRAC), has published enforcement data gathered via FOIA from the Bureau of Alcohol, Tobacco, and Firearms (ATF), the Department of Homeland Security (DHS), the Department of Justice (DOJ), the Drug Enforcement Administration (DEA), the Federal Bureau of Investigation (FBI), and the Internal Revenue Service (IRS).¹⁴³

137. Schauer, *supra* note 9, at 1347; Allen Ferrell, *The Case for Mandatory Disclosure in Securities Regulation Around the World*, 2 BROOK. J. CORP. FIN. & COM. L. 81 (2007); Edward Rock, *Securities Regulation As Lobster Trap: A Credible Commitment Theory of Mandatory Disclosure*, 23 CARDOZO L. REV. 675 (2002); Sage, *supra* note 5, at 1780 (noting that the “SEC penalizes faulty disclosure, not faulty performance”).

138. Michael B. Rothberg et al., *Choosing the Best Hospital: The Limitations of Public Quality Reporting*, 27 HEALTH AFF. 1680 (2008); Schauer, *supra* note 9, at 1348.

139. Jeremy Bentham, *Farming Defended*, in 1 WRITINGS ON THE POOR LAWS 276, 277 (Michael Quinn ed., Oxford Univ. Press 2001 (1796)); Schauer, *supra* note 9, at 1352.

140. See discussion *infra* Part IV, regarding several examples from the EPA, FEC, FDA, etc.

141. Vladeck, *supra* note 64, at 1830–31.

142. *Id.* at 1830.

143. *Id.*; TRAC: *About Us*, TRANSACTIONAL RECORDS ACCESS CLEARINGHOUSE, <http://trac.syr.edu/aboutTRACgeneral.html> (last visited July 15, 2017) [<https://perma.cc/L9NF-VZ4R>]. The FTC maintains a massive database of complaints against companies, though it is nonpublic and is available only to enforcement agencies, such as the FTC, Department of Justice, and participating state and local agencies. *Consumer Sentinel Network*, FED. TRADE COMM’N, <https://www.ftc.gov/enforcement/consumer-sentinel-network> (last visited July 15,

Enforcement data may also help counter underenforcement by agencies, what Matthew Stephenson calls “agency slack.”¹⁴⁴ Scholars note widespread underenforcement by a variety of agencies in a variety of contexts.¹⁴⁵ Even when agencies do pursue regulatory violations, they often fail to enforce them.¹⁴⁶ Underenforcement may derive from several sources—insufficient agency resources, ideology, anti-regulatory pressures, political oversight, inertia, or agency self-interest.¹⁴⁷ Regardless of the contributors, making compliance and enforcement data public might inspire agencies to reach more optimal levels of enforcement—or even inspire companies to reach more optimal levels of compliance. Observers also suspect the converse. When the U.S. Department of Agriculture (USDA) “abruptly removed inspection reports, warning letters, and other documents on nearly 8000 animal facilities that the agency regulates” from its website in February 2017, critics worried that the move would shield violators from journalists and animal rights groups.¹⁴⁸

If publication alone does not encourage compliance, perhaps use of the data by third-party intermediaries can. In justifying its Consumer Complaint Database, the CFPB pointed to companies offering commercial intelligence products based on data from the FDA’s drug and device adverse events databases (FAERS and MAUDE).¹⁴⁹ The CFPB itself notes that

2017) [<https://perma.cc/6Z24-SBVJ>]; see NATHAN CORTEZ, ADMIN. CONFERENCE OF THE U.S., AGENCY PUBLICITY IN THE INTERNET ERA (2015) <https://www.acus.gov/sites/default/files/documents/agency-publicity-in-the-internet-era.pdf> [<https://perma.cc/6Z24-SBVJ>].

144. Matthew C. Stephenson, *Public Regulation of Private Enforcement: The Case for Expanding the Role of Administrative Agencies*, 91 VA. L. REV. 93, 110 (2005).

145. See Nathan Cortez, *Regulating Disruptive Innovation*, 29 BERKELEY TECH. L.J. 173, 220–21 (2014); Stephen Lee, *Private Immigration Screening in the Workplace*, 61 STAN. L. REV. 1103, 1126–29 (2009); Stephenson, *supra* note 144, at 116–17; Rena Steinzor, *The Truth About Regulation in America*, 5 HARV. L. & POL’Y REV. 323 (2011).

146. See, e.g., Ezra Ross & Martin Pritikin, *The Collection Gap: Underenforcement of Corporate and White-Collar Fines and Penalties*, 29 YALE L. & POL’Y REV. 453, 473–74 (2011).

147. Stephenson, *supra* note 144.

148. Meredith Wadman, *Courts Ponder How Public Animal Reports Must Be*, 356 SCIENCE 790, 790 (2017).

149. Disclosure of Certain Credit Card Complaint Data, 77 Fed. Reg. 37,558, 37,562 (June 22, 2012) (citing Melinda Beck, *Searching for Side Effects*, WALL ST. J.: HEALTH J. (Jan. 31, 2012), <https://www.wsj.com/articles/SB1000142405297020>

third-party users, like the consulting firm Deloitte and the *U.S. News & World Report*, are relying on the Bureau's data to publish findings and recommendations.¹⁵⁰ Moreover, there is always the lingering fear that shareholders, plaintiffs' lawyers, media, bloggers, or other enforcement agencies will use published data against companies.

An emerging potential use of government data, and one encouraged by the government itself, is "crowdsourcing." Crowdsourcing is a method of soliciting answers, ideas, resources, or services from a large network of people, typically online.¹⁵¹ Technologists have envisioned the government serving as a "platform" for innovation by providing data that inspires outside parties to create innovative uses for the data.¹⁵² Government agencies cannot predict how their data sets might be used by the public, but the act of publishing data in raw, open, and machine-readable format allows the public to generate innovative and perhaps more enlightening uses of the data.¹⁵³

The Obama administration pursued several crowdsourcing initiatives, published on websites like Challenge.gov, which features various prize competitions sponsored by over eighty federal agencies.¹⁵⁴ The site claims that the government has awarded "[m]ore than \$250 million in prize money" since 2010.¹⁵⁵ The idea is that "U.S. federal agencies invite the public's help to solve perplexing mission-centric problems."¹⁵⁶ Indeed, even the administration's original Open Government Directive offered contests and prizes that incentivized the

3920204577193052426275904 [https://perma.cc/872W-TUP2] (referring to products by AdverseEvents, Inc. and Clarimed, LLC)).

150. CORTEZ, *supra* note 143, at 66.

151. *Crowdsourcing*, MERRIAM-WEBSTER DICTIONARY, <http://www.merriam-webster.com/dictionary/crowdsourcing> (last visited July 15, 2017) [https://perma.cc/YB4X-TV3S].

152. *See, e.g.*, Tim O'Reilly, *Government as a Platform*, in OPEN GOVERNMENT: COLLABORATION, TRANSPARENCY, AND PARTICIPATION IN PRACTICE 11 (Daniel Lathrop & Laurel Ruma eds., 2010); Shkabatur, *supra* note 55, at 110.

153. Shkabatur, *supra* note 55, at 109.

154. *About*, CHALLENGE.GOV, <http://www.challenge.gov/about/> (last visited July 15, 2017) [https://perma.cc/ENV8-9HCT].

155. *Id.*

156. *Id.*

public to “tinker” with the data released.¹⁵⁷ Thus, some of the most provocative uses of agency data may just be emerging.

C. *Democratic Ideals*

A third and more lofty justification for publishing government data is to enhance government accountability.¹⁵⁸ Transparency is often assumed to be its precondition.¹⁵⁹ Leading thinkers—such as John Milton, John Stuart Mill, James Madison, Oliver Wendell Holmes (father of the marketplace of ideas),¹⁶⁰ and Louis Brandeis—have long drawn an explicit link between transparency, accountability, and democracy.¹⁶¹ Centuries of writing are filled with paeans to the virtues of transparency. Today, the modern open source movement, which owes much to these forebears and helped seed the open government movement, holds that information is a necessary precondition for truth and progress.¹⁶² Seen in this way, agency databases are a logical vehicle for both technological and democratic progress.

157. Exec. Office of the President, Memorandum for the Heads of Executive Departments and Agencies, Ser. No. M-10-11 (2010); Shkabatur, *supra* note 55, at 111 n.173.

158. Shkabatur, *supra* note 55, at 80; Adam M. Samaha, *Government Secrets, Constitutional Law, and Platforms for Judicial Intervention*, 53 UCLA L. REV. 909, 917 (2006).

159. Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government (Jan. 21, 2009), 74 Fed. Reg. 4685, 4685 (Jan. 21, 2009); Fenster, *supra* note 26, at 894–99; Shkabatur, *supra* note 55, at 83.

160. *Abrams v. United States*, 250 U.S. 616, 630 (1919) (Holmes, J., dissenting).

161. See, e.g., Letter from James Madison to W.T. Barry (Aug. 4, 1822), in 9 THE WRITINGS OF JAMES MADISON 103 (Gaillard Hunt ed., 1910) (“A popular Government, without popular information, or the means of acquiring it, is but a Prologue to a Farce or a Tragedy; or perhaps both.” As FUNG ET AL., *supra* note 15, at 24, discuss, this language is carved on the Library of Congress building); *Abrams*, 250 U.S. at 630 (Holmes, J., dissenting) (“[T]he best test of truth is the power of [an assertion] to get itself accepted in the competition of the market . . .”); Stanley Ingber, *The Marketplace of Ideas: A Legitimizing Myth*, 1984 DUKE L.J. 1, 2–3 (1984).

162. Schauer, *supra* note 9, at 1350. However, despite the fanfare, some scholars question whether recent transparency and open government initiatives actually “strengthen public accountability.” See, e.g., Shkabatur, *supra* note 55, at 81. Shkabatur argues that current transparency policies, driven by new technologies, give agencies too much control over which data are published, “prioritizes quantity over quality of disclosures,” and reinforces older barriers to accessing information. *Id.*

Government databases can also serve an expressive function by acting as a conduit for consumers to air their grievances.¹⁶³ Presidents Kennedy and Nixon both promulgated a Consumer Bill of Rights to pose the government as an intermediary or a tribunal through which consumer complaints could be given a voice.¹⁶⁴ Such a role can increase the public's confidence in government, providing "a positive point of contact" between agencies and citizens, and promoting the idea of "government as a positive force in society."¹⁶⁵ The CFPB's Consumer Complaint Database essentially serves this role today.

In short, there are compelling reasons why data disclosure appeals to policymakers so much. Disclosure seems consistent with free-market and autonomy principles, and seems to be an easy and effective intervention compared to more traditional regulation.¹⁶⁶ Politically, regulation by disclosure is cast as a "path of least resistance for administrative agencies seeking to promote meaningful change."¹⁶⁷ Indeed, calls for disclosure and transparency are often justified, on a more fundamental plane, as furthering the pursuit of truth, knowledge, and societal progress.¹⁶⁸

III. DATA AND ITS DISCONTENTS

But does disclosure live up to its many promises? In this Part, I consider the shortcomings of database disclosures, before evaluating several prominent agency databases in Part IV. Part V then considers ways to design databases for more optimal, effective disclosure.

163. Katherine Porter, *The Complaint Conundrum: Thoughts on the CFPB's Complaint Mechanism*, 7 BROOK. J. CORP. FIN. & COM. L. 57, 75–76 (2012). Again, there is a large literature on the expressive function of disclosure regimes. See, e.g., Cass R. Sunstein, *On the Expressive Function of Law*, 144 U. PA. L. REV. 2021 (1996). *Contra* Matthew D. Adler, *Expressive Theories of Law: A Skeptical Overview*, 148 U. PA. L. REV. 1363 (2000).

164. Porter, *supra* note 163, at 76; Special Message to the Congress on Protecting the Consumer Interest, 1962 PUB. PAPERS 235, 236 (Mar. 15, 1962); Special Message to the Congress on Consumer Protection, 1969 PUB. PAPERS 883, 883 (Oct. 30, 1969).

165. Porter, *supra* note 163, at 76.

166. Ben-Shahar & Schneider, *supra* note 17, at 681–82.

167. Sage, *supra* note 5, at 1772.

168. Schauer, *supra* note 9, at 1350.

First and foremost, agency disclosure of information about regulated parties can itself cause a variety of harms, ranging from concrete (a devaluation of stock price) to less tangible, reputational harms.¹⁶⁹ Although scholarship on these harms is not voluminous, the harms are relatively well documented.¹⁷⁰ For example, in 2008 the FDA and CDC mistakenly identified tomatoes as the source of a salmonella outbreak, costing the tomato industry an estimated \$200 million in lost sales.¹⁷¹ And there are numerous instances in which a company's stock value plummeted after an agency announcement criticized the company.¹⁷² As Vladeck notes, "[t]here is also force, as a general matter, to the argument that companies should not be subject to commercial harm simply because they are compelled to report their activities to the government."¹⁷³

Questions about the value of so-called "naming and shaming" have crept into various disciplines.¹⁷⁴ In the book *Is Shame Necessary?*, Jennifer Jacquet considers the virtues and flaws of modern naming and shaming, particularly how social media and other modern modes of communication might amplify shame effectively to change the behavior of corporations or even governments.¹⁷⁵ Indeed, organizations like Wikileaks derive tremendous power and influence from disclosure—the kind of nongovernmental power normally reserved for mainstream media.¹⁷⁶

169. See, e.g., Ernest Gellhorn, *Adverse Publicity by Administrative Agencies*, 86 HARV. L. REV. 1380 (1973); Cortez, *supra* note 75; CORTEZ, *supra* note 143, at 9–12.

170. For just a few examples, see James T. O'Reilly, *The 411 on 515: How OIRA's Expanded Information Roles in 2002 Will Impact Rulemaking and Agency Publicity Actions*, 54 ADMIN. L. REV. 835 (2002); O'Reilly, *supra* note 83; Lars Noah, *Administrative Arm-Twisting in the Shadow of Congressional Delegations of Authority*, 1997 WIS. L. REV. 874 (1997); Wayne A. Logan & Andrew Guthrie Ferguson, *Policing Criminal Justice Data*, 101 MINN. L. REV. 541 (2016).

171. Denis G. Maki, *Coming to Grips with Foodborne Infection—Peanut Butter, Peppers, and Nationwide Salmonella Outbreaks*, 360 NEW ENG. J. MED. 949 (2009).

172. See Cortez, *supra* note 75 (discussing examples); see also CORTEZ, *supra* note 143, at 9–12.

173. Vladeck, *supra* note 64, at 1793 (noting, however, that secrecy should give way to publication, for example, when necessary to avoid death or serious injury).

174. See, e.g., BRENT FISSE & JOHN BRAITHWAITE, *THE IMPACT OF PUBLICITY ON CORPORATE OFFENDERS* (1983).

175. JENNIFER JACQUET, *IS SHAME NECESSARY: NEW USES FOR AN OLD TOOL* (2015).

176. Shkabatur, *supra* note 55, at 135–36.

Less sanguine views argue that “shaming is the very antithesis of the law,”¹⁷⁷ particularly when wielded by the government. For example, Donald Trump’s Twitter posts, the focus of significant handwringing, demonstrate the unfair destructive power of adverse publicity. In December 2016, Boeing stock took a quick plunge after President-elect Trump took to Twitter to criticize the cost of Boeing’s contract to build a new Air Force One.¹⁷⁸ After the episode, some investors and market analysts began to monitor Trump’s tweets for potential market-moving proclamations.¹⁷⁹

A second objection to the use of disclosure as a regulatory tool is that it may be ineffective. There is growing scholarly skepticism that openness necessarily leads to knowledge or that more information necessarily produces better decisions.¹⁸⁰ As David Vladeck observes, “there is now a significant and growing dissonance between the promises made by our federal right-to-know laws and their performance.”¹⁸¹ And as Ben-Shahar and Schneider emphasize in their magisterial article, *The Failure of Mandated Disclosure*, disclosure regimes often fail completely in meeting their goals, and in fact can have unintended consequences that hinder them.¹⁸² Although their bearish views are challenged,¹⁸³ the virtues of disclosure no longer remain uncontested.

A third objection to agency disclosure in general, and to databases in particular, is that they are not always complete or accurate. To be most useful, data that purport to present truthful, objective information must meet those standards.

177. Eric Posner, *A Terrible Shame: Enforcing Moral Norms Without Law Is No Way to Create a Virtuous Society*, SLATE (Apr. 9, 2015, 11:14 A.M.), http://www.slate.com/articles/news_and_politics/view_from_chicago/2015/04/inter_net_shaming_the_legal_history_of_shame_and_its_costs_and_benefits.html [https://perma.cc/TT6K-GEB5].

178. Bump, *supra* note 22.

179. Alyssa Sims, *The Stock Market Needs to Gird Itself for Trump’s Tweets*, SLATE (Dec. 16, 2016), http://www.slate.com/articles/technology/future_tense/2016/12/the_stock_market_needs_to_gird_itself_for_trump_s_tweets.html [https://perma.cc/N6FU-V7ZF].

180. Schauer, *supra* note 9, at 1351.

181. Vladeck, *supra* note 64, at 1792.

182. Ben-Shahar & Schneider, *supra* note 17.

183. Craswell, *supra* note 18; Ryan Bubb, *TMI? Why the Optimal Architecture of Disclosure Remains TBD*, 113 MICH. L. REV. 1021 (2015); Kristin Madison, *Health Care Quality Reporting: A Failed Form of Mandated Disclosure?*, 13 IND. HEALTH L. REV. 310 (2016).

Otherwise, they risk succumbing to a problem known well to computer and software engineers: “Garbage in, garbage out.”¹⁸⁴ Flawed inputs produce flawed outputs, and inaccurate databases will be unreliable. Even well-known advocates of regulation by information, such as Cass Sunstein, acknowledge the dangers of regulating based on flawed data.¹⁸⁵ Litigation challenging inaccurate reports on the CPSC’s SaferProducts.gov database demonstrates that agencies sometimes post inaccurate data, and sometimes that data lingers online for years before it is corrected or retracted.¹⁸⁶ The FEC’s campaign finance data is widely acknowledged to be incomplete.¹⁸⁷ The FDA’s medical device database, which tracks problems with devices and is subject to mandatory reporting requirements, is undermined dramatically by under-reporting.¹⁸⁸ Even relatively noncontroversial databases like USAspending.gov and Recovery.gov have been found to have widespread inaccuracies.¹⁸⁹ Both the OMB and GAO have acknowledged that data on the sites has been inaccurate, untimely, or incomplete.¹⁹⁰ An independent review by the Sunlight Foundation found that the sites had “over 1.2 trillion dollars’ worth of misreported spending in 2009 alone.”¹⁹¹

184. *Garbage in, garbage out*, OXFORD LIVING DICTIONARIES, https://en.oxforddictionaries.com/definition/us/garbage_in_garbage_out (last visited July 15, 2017) [<https://perma.cc/LF2P-4VDK>].

185. Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683, 755–60 (1999).

186. *See, e.g.*, *Co. Doe v. Tenenbaum*, 900 F. Supp. 2d 572 (D. Md. 2012), *rev’d sub nom.* *Co. Doe v. Public Citizen*, 749 F.3d 246 (4th Cir. 2014).

187. *See* discussion *infra* Section IV.A.

188. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO/HEHS-97-21, MEDICAL DEVICE REPORTING: IMPROVEMENTS NEEDED IN FDA’S SYSTEM FOR MONITORING PROBLEMS WITH APPROVED DEVICES (1997), <http://www.gao.gov/products/HEHS-97-21> [<https://perma.cc/65CQ-WF8W>].

189. *See, e.g.*, Shkabatur, *supra* note 55, at 103.

190. OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, M-09-21, IMPLEMENTING GUIDANCE FOR THE REPORTS ON USE OF FUNDS PURSUANT TO THE AMERICAN RECOVERY AND REINVESTMENT ACT OF 2009, at 27 (2009), <https://ojp.gov/recovery/pdfs/arrafundsmemo.pdf> [<https://perma.cc/5AZL-FLRY>]; Memorandum from Robert Shea, Assoc. Dir., Office of Mgmt. & Budget, to Fed. Agencies 3 (Mar. 6, 2008) (on file with author).

191. *Transparency Through Technology: Evaluating Federal Open-Government Initiatives: Hearing Before the Subcomm. on Tech., Info. Policy, Intergovernmental Relations, and Procurement Reform of the H. Comm. on Oversight and Gov’t Reform*, 112th Cong. 3 (2011) (statement of Ellen Miller, Executive Director, Sunlight Foundation); Shkabatur, *supra* note 55, at 103.

Inaccurate or incomplete data sets can be even more problematic if agencies process them with software algorithms or artificial intelligence to identify regulatory violators or single out firms for further investigation.¹⁹² The use of such algorithms for regulatory enforcement purposes raises a host of novel questions about agency delegations, justifications, and reasoning.¹⁹³ Moreover, increased skepticism of the accuracy and objectivity of algorithmic decision making warrants further investigation, particularly as regulators rely on these methods more.¹⁹⁴

A related criticism of agency databases is that they often present data without appropriate context.¹⁹⁵ Jennifer Shkabatur notes that it is hard for a lay person browsing USASpending.gov, for example, to evaluate whether a \$20 million contract between the Department of Commerce and Industrial Economics for “continued support for the Deepwater Horizon oil spill” is money well-spent, or whether an \$817 million contract between the Department of Defense and Lockheed Martin for “incremental funding” is wasteful.¹⁹⁶ But that is the only information provided. Thus, she argues, “even if data is timely and reliable,” when stripped of context, it may not always be particularly meaningful or useful.¹⁹⁷ Providing appropriate context is one way for agencies to act as data

192. For example, the FTC uses data from its nonpublic Sentinel database to look for patterns of complaints against a potential enforcement target. See CORTEZ, *supra* note 143, at 55–56. For a series of posts on the use of automated decision-making by regulatory agencies, see Mariano-Florentino Cuéllar, *Artificial Intelligence and the Administrative State*, REG. REVIEW (Dec. 19–22, 2016), <https://www.theregreview.org/2016/12/19/artificial-intelligence-and-the-administrative-state/> [<https://perma.cc/WS9U-USFN>].

193. See, e.g., Mariano-Florentino Cuéllar, *Cyberdelegation and the Administrative State*, in ADMINISTRATIVE LAW FROM THE INSIDE OUT: ESSAYS ON THEMES IN THE WORK OF JERRY L. MASHAW 134 (Nicholas R. Parrilo ed., 2017).

194. See *id.*; see also FRANK PASQUALE, *THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION* (2015) (investigating algorithms in the private sector); CATHY O’NEIL, *WEAPONS OF MATH DESTRUCTION: HOW BIG DATA INCREASES INEQUALITY AND THREATENS DEMOCRACY* (2016) (questioning the objectivity of algorithms and machine-learning tools).

195. See, e.g., Shkabatur, *supra* note 55, at 104 (citing reports and concerns that federal spending databases, like USASpending.gov and Recovery.gov, provide voluminous data on spending, but do not provide criteria or justifications for those spending decisions and whether they square with congressional directives).

196. *Id.* at 104–05 (noting that these are two real-life examples in USASpending.gov).

197. *Id.* at 105.

“stewards,” rather than to serve as mere publishers or repositories.

Scholars also note that agencies exercise significant discretion to decide what data to disclose and the scope of that data, which can skew user perceptions.¹⁹⁸ Industry repeatedly made such arguments as the CFPB built its Consumer Complaint Database. Commenters objected that the database would necessarily include only self-selected complaints that were nonrandom and thus not representative of the consumer population.¹⁹⁹ Industry commenters also objected that the complaints lacked context and that users might overlook the data’s limitations, despite disclaimers by the Bureau.²⁰⁰

Another problem with agency disclosure is the volume of it, ironically, and the risk of drowning the public with information. Data.gov currently hosts over 195,000 data sets,²⁰¹ some of which include millions of records or data points. But more information is not always better information—although many disclosure regimes assume otherwise.²⁰² Today, federal agencies post so much information online that many scholars wonder who is served by these data dumps.²⁰³ The conventional wisdom is that few benefit from an “undifferentiated mass of information” posted online, as the “cost of sifting through it would overwhelm its value.”²⁰⁴ Genuinely useful information is often buried.²⁰⁵ As Paredes notes, sunlight can be the best disinfectant, “[b]ut sunlight can also be blinding.”²⁰⁶

So how does salient information stand out? Evaluations of mandatory disclosures regimes find that even the most well-known are ignored: “Next to the warning label on cigarette packs, *Miranda* is the most widely ignored piece of official advice in our society.”²⁰⁷ Gradually, because mandated

198. *Id.* at 117.

199. *See, e.g.*, Disclosure of Certain Credit Card Complaint Data, 77 Fed. Reg. 37,558, 37,561 (June 22, 2012).

200. *See id.* at 37,562.

201. DATA.GOV, <http://www.data.gov> (last visited Feb. 14, 2017) [<https://perma.cc/E3N3-NPSG>].

202. *See* Ben-Shahar & Schneider, *supra* note 17, at 650.

203. Shkabatur, *supra* note 55, at 118.

204. Vladeck, *supra* note 64, at 1832.

205. Ben-Shahar & Schneider, *supra* note 17, at 737.

206. Paredes, *supra* note 10, at 419.

207. Richard A. Leo, *Questioning the Relevance of Miranda in the Twenty-First Century*, 99 MICH. L. REV. 1000, 1012–13 (quoting Patrick Malone, “*You Have the*

disclosures are so attractive to policymakers, they tend to accumulate over time, which only compounds the problem—“disclosures are added, never removed.”²⁰⁸ As Ben-Shahar and Schneider observe, consumers “encounter too many disclosures to digest most of them.”²⁰⁹ Such concerns lend credence to the idea that sophisticated intermediaries will continue to serve a valuable channeling and interpretive role. Indeed, the idea behind the massive data dumps on Data.gov and other federal databases “is that nongovernmental intermediaries can step in and translate the raw data for the general public.”²¹⁰

A related shortcoming of disclosure is the complexity of the data. Mandated disclosure regimes have become ubiquitous in federal and state law, ensconced in statutes, ordinances, agency regulations, and common law.²¹¹ Sometimes, these sources of law demand “marvelously elaborate disclosures” that are difficult if not impossible for the intended beneficiaries (usually consumers) to understand.²¹² Consumers, of course, are not perfectly rational, but exercise “bounded rationality” due to various cognitive biases and distortions.²¹³ Thus, many scholars have come to recognize that disclosure of objective information may not, in itself, generate optimal outcomes—rather, disclosure regimes “may need to aggregate, translate, simplify, or benchmark the facts.”²¹⁴ Even if users do not understand the science or statistical techniques behind the data, the data may still be successful in improving the product or conduct targeted by the disclosure.²¹⁵

Given the volume and complexity of most data, intended beneficiaries often do not and cannot use it, particularly in the

Right to Remain Silent: Miranda After Twenty Years, 55 AM. SCHOLAR 367, 368 (1986); Ben-Shahar & Schneider, *supra* note 17, at 678.

208. Ben-Shahar & Schneider, *supra* note 17, at 679.

209. *Id.* at 705.

210. Shkabatur, *supra* note 55, at 118.

211. Ben-Shahar & Schneider, *supra* note 17, at 650.

212. *Id.*

213. FUNG ET AL., *supra* note 15, at 33. This is in contrast to the perfectly rational “economic man” famously described by Herbert Simon, which he recognized is an ideal rather than a reality. Herbert A. Simon, *A Behavioral Model of Rational Choice*, 69 Q.J. ECON. 99, 99 (1955). See generally HERBERT A. SIMON, *MODELS OF BOUNDED RATIONALITY: ECONOMIC ANALYSIS AND PUBLIC POLICY* (1982).

214. FUNG ET AL., *supra* note 15, at 34.

215. Craswell, *supra* note 18, at 361.

idealized way policymakers intend. In a variety of legal contexts, the targets for information disclosure “often do not read disclosed information, do not understand it when they read it, and do not use it even if they understand it.”²¹⁶ Examples abound. A troubling one is the extensive campus crime data reported by colleges and universities to the Department of Education under the Clery Act,²¹⁷ which often goes unread.²¹⁸ Nevertheless, the Department promises that the Clery Act “is intended to provide students and their families, as higher education consumers, with accurate, complete, and timely information about safety on campus so that they can make informed decisions.”²¹⁹ The current reality is that, for most databases, such aspirations outstrip reality.

Databases and other disclosure regimes can also be costly. Disclosure is often assumed to be simple and low-cost, particularly compared to more conventional regulation and enforcement.²²⁰ But successful disclosure systems often require “a distinctive and demanding architecture.”²²¹ Any thoughtful disclosure regime must determine what information must be disclosed, by whom, to whom, the optimal format, and appropriate quality assurance and enforcement mechanisms.²²² Indeed, many assume that disclosure avoids many of the compliance and enforcement costs that attend to traditional command-and-control regulation.²²³ But, compliance with disclosure regimes must also be monitored and enforced—frequently with both civil and criminal penalties.²²⁴ For regimes with low rates of compliance, policymakers have often increased both penalties and enforcement efforts.²²⁵

216. Ben-Shahar & Schneider, *supra* note 17, at 665.

217. Disclosure of Campus Security Policy and Campus Crime Statistics (Clery) Act, Pub. L. No. 101-542, 104 Stat. 2385 (1990) (codified at 20 U.S.C. § 1092(f)).

218. Ben-Shahar & Schneider, *supra* note 17, at 702–04; Sara Lipka, *Do Crime Statistics Keep Students Safe?*, CHRON. HIGHER EDUC., Jan. 30, 2009, at A1.

219. OFFICE OF POSTSECONDARY EDUC., U.S. DEP’T OF EDUC., *THE HANDBOOK FOR CAMPUS CRIME REPORTING* 3 (2005); Ben-Shahar & Schneider, *supra* note 17, at 702–03.

220. Ben-Shahar & Schneider, *supra* note 17, at 735–37.

221. FUNG ET AL., *supra* note 15, at 39.

222. *Id.* at 39–46.

223. Ben-Shahar & Schneider, *supra* note 17, at 735–37.

224. FUNG ET AL., *supra* note 15, at 45 (examining civil and criminal penalties to enforce the disclosure mandates in campaign finance law and corporate law).

225. *Id.* at 45–46.

Moreover, the cost to disclosers can be significant. For example, a single new SEC requirement that companies file “current reports” of insider transactions was expected to generate 215,000 additional filings to the SEC annually, at an estimated cost of over \$89 million per year.²²⁶ Another recent study found that U.S. physicians in just four common specialties spend \$15.4 billion annually reporting under various quality measurement programs.²²⁷ Thus, regulation by disclosure can be costly.

Opportunity costs can also be significant. Relying on disclosure as a means to pursue regulatory ends may mean bypassing other, better means for achieving those ends.²²⁸ Traditional command-and-control regulation long ago lost its luster among policymakers and academics.²²⁹ But modern replacements like “new governance,” despite their many promises, can underwhelm. Despite the widespread use of disclosure, “it remains an open question whether transparency as regulation is better or worse, all things considered, than more direct forms of regulation.”²³⁰ One might reasonably wonder whether the time and personnel that agencies spend on disclosure would be better spent writing regulations and enforcing them.

Finally, there is some evidence that mandatory disclosure regimes can backfire. Consumers might be tempted to let their guards down when presented with mandated disclosures that give transactions a “veneer of legality.”²³¹ There is also

226. Form 8-K Disclosure of Certain Management Transactions, 67 Fed. Reg. 19,914 (proposed Apr. 23, 2002) (to be codified at 17 CFR pts. 230, 239, 249).

227. Lawrence B. Casalino et al., *U.S. Physician Practices Spend More Than \$15.4 Billion Annually to Report Quality Measures*, 35 HEALTH AFF. 401 (Mar. 2016). Of course, for perspective, annual U.S. health spending is roughly \$3.2 trillion. *Historical*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/national-healthexpenddata/nationalhealthaccountshistorical.html> (last visited July 14, 2017) [<http://perma.cc/B5YW-ZZAJ>].

228. See Ben-Shahar & Schneider, *supra* note 17, at 737–42.

229. Disclosure also responds to the Hayekian criticism that command-and-control regulation is insufficient for regulating a large, complex society. F.A. Hayek, *The Use of Knowledge in Society*, 35 AM. ECON. REV. 519, 524–25 (1945).

230. Schauer, *supra* note 9, at 1348.

231. Ben-Shahar & Schneider, *supra* note 17, at 740 (citing Lauren E. Willis, *Decisionmaking and the Limits of Disclosure: The Problem of Predatory Lending: Price*, 65 MD. L. REV. 707, 794–95 (2006) (using as an example the many disclosures required in consumer loans, which give a “veneer of legality and authority to the loan process”)).

evidence that the party required to disclose information often interprets their compliance with the disclosure requirement as granting them license to act more harshly.²³² Although there is much more research to be done, again, the virtues of disclosure are no longer uncontested.

IV. AGENCY DATABASES

Nascent skepticism with the use of disclosure as a regulatory tool has not stopped disclosure efforts from proliferating. Today, thousands of federal agency websites host hundreds of thousands of agency databases (as of February 2017, over 195,000).²³³ As such, writing about government data is difficult because databases have become so very common.²³⁴ A comprehensive survey of agency databases would “pointlessly burden” the audience, risking the same information overload often imposed by policymakers that rely on disclosure.²³⁵ Nevertheless, as becomes quickly obvious,²³⁶ agency databases are becoming the norm rather than the exception.

Below I survey six of the most salient agency efforts to post searchable data sets online. Of course, there are many more worth discussion—too many for a single article. The following represents a cross-section of databases that try to achieve regulatory goals, using online disclosure of behavior to try to affect that behavior.

A. *The FEC’s Campaign Finance Data*

For decades, federal campaign finance law has relied on disclosure not only to police limits on campaign contributions and spending, but also to pursue deeper goals of deterring

232. *Id.* at 739 (citing such evidence related to conflict-of-interest disclosures).

233. DATA.GOV, *supra* note 201 (“Search over 194,126 data sets” as of March 22, 2016). Note that as of April 1, 2016, 158,301 data sets were published by the federal government, with the rest published by sub-federal units of government, including a smattering from the private and nonprofit sectors. *Data Catalog: Organization Types*, DATA.GOV, http://catalog.data.gov/dataset#sec-organization_type (last visited July 14, 2017) [<http://perma.cc/L5SY-5M9S>] (See “Organization Types” in the left sidebar).

234. Ben-Shahar & Schneider, *supra* note 17, at 652.

235. *Id.*

236. CORTEZ, *supra* note 143, at app. E.

corruption and the appearance of it.²³⁷ Indeed, disclosure has been, perhaps, the one leg of the campaign finance law tripod to be spared by the Supreme Court.²³⁸ Unlike limits on campaign contributions and expenditures, disclosure requirements have endured repeated First Amendment challenges—from early cases like *Burroughs*²³⁹ to more strident recent cases like *Citizens United* and *McCutcheon*.²⁴⁰ Of course, in the seminal case *Buckley v. Valeo*,²⁴¹ the Court upheld various disclosure requirements introduced in the Federal Election Campaign Act of 1971 (FECA) and the FECA Amendments of 1974.²⁴² In the ensuing decades, as courts invalidated various restrictions on campaign contributions and expenditures on First Amendment grounds, disclosure requirements endured. Thus, through attrition, disclosure has become the preferred choice for regulating money in politics.²⁴³ Indeed, recent reform proposals would address lingering problems with campaign finance regulation by using even more disclosures, for example, by trying to triangulate FEC data with data from other agencies like the SEC and IRS.²⁴⁴

Today, the FEC maintains several searchable online databases on its website, which are centralized on the FEC's Campaign Finance Disclosure Portal.²⁴⁵ Users can search FEC data based on reports required of federal candidates, parties,

237. *Buckley v. Valeo*, 424 U.S. 1, 66–68 (1976).

238. The other two legs being limits on campaign contributions and limits on expenditures.

239. *Burroughs v. United States*, 290 U.S. 534 (1934) (upholding an indictment under the Federal Corrupt Practices Act).

240. *Citizens United v. Fed. Election Comm'n*, 558 U.S. 310, 319 (2010); *McCutcheon v. Fed. Election Comm'n*, 134 S. Ct. 1434, 1442, 1459–60 (2014).

241. *Buckley*, 424 U.S. at 66–68.

242. Federal Election Campaign Act of 1971 (FECA), Pub. L. No. 92-225, 86 Stat. 3 (1972), amended by Federal Election Campaign Act Amendments of 1974, Pub. L. No. 93-443, 88 Stat. 1263 (codified as amended at 2 U.S.C. § 431).

243. See Jennifer A. Heerwig & Katherine Shaw, *Through a Glass Darkly: The Rhetoric and Reality of Campaign Finance Disclosure*, 102 GEO. L.J. 1443, 1445 (2014); Richard L. Hasen, *Chill Out: A Qualified Defense of Campaign Finance Disclosure Laws in the Internet Age*, 27 J.L. & POL. 557, 572 (2012).

244. See, e.g., Lucian A. Bebchuk & Robert J. Jackson, Jr., *Shining Light on Corporate Political Spending*, 101 GEO. L.J. 923, 931, 935–36 (2013).

245. *Campaign Finance Disclosure Portal*, FED. ELECTION COMM'N, <http://www.fec.gov/pindex.shtml> (last visited July 14, 2017) [<http://perma.cc/V7V5-Q4MR>] [hereinafter FEC]. The FEC is currently transitioning to a new website. See Press Release, Fed. Election Comm'n, FEC introduces retooled, streamlined, user-centered website (May 30, 2017), <https://www.fec.gov/updates/fec-introduces-retooled-streamlined-user-centered-website/> [<https://perma.cc/U8N8-ADXZ>].

committees, donors, and lobbyists, among others.²⁴⁶ The data are generally searchable by name, date, and location,²⁴⁷ and are presented in list, map, and chart form,²⁴⁸ making them more accessible to lay users. Thus, for example, a user can learn how much money a federal candidate in her district has raised and spent, or how much a certain political action committee (PAC) has dedicated to electioneering, or find detailed information about independent expenditures or bundled contributions.²⁴⁹ The FEC's Disclosure Data Catalog publishes these datasets in downloadable .CSV, .XML, or .XSD formats,²⁵⁰ thus making them more useful for sophisticated users.

Despite their broad scope, FEC databases pay special attention, of course, to the activities of candidates and committees. Users can search for federal candidates and political committees by name and view on a single page all reports filed by that person or committee, including a multi-year summary of the money they raised and spent.²⁵¹

Like many other agencies, the FEC also publishes online searchable databases of enforcement records. Its Enforcement Query System is a searchable depository of FEC enforcement documents, including complaints, responses, settlements, and other relevant documents.²⁵² The system includes, for example, a searchable and sortable list of parties required to pay

246. *Campaign Finance Reports and Data*, FED. ELECTION COMM'N, <https://www.fec.gov/data/> (last visited July 22, 2017) [<https://perma.cc/5BB9-FH3X>].

247. *Searchable Presentations*, FED. ELECTION COMM'N, <http://www.fec.gov/portal/searchable.shtml> (last visited July 22, 2017) [<https://perma.cc/T6TN-RGYN>].

248. *See, e.g., 2016 Presidential Campaign Finance*, FED. ELECTION COMM'N, <http://www.fec.gov/disclosure/pnational.do> (last visited July 22, 2017) [<https://perma.cc/DL4N-VCUF>].

249. *Disclosure Data Catalog*, FED. ELECTION COMM'N, <http://www.fec.gov/data/DataCatalog.do?cf=searchable> (last visited July 22, 2017) [<https://perma.cc/V6FV-KMME>].

250. *Id.*

251. *See, e.g., Financial Summary for Ted Cruz, Full Cycle: 2013-18*, FED. ELECTION COMM'N, <https://www.fec.gov/data/candidate/S2TX00312/> (last visited Aug. 21, 2017) [<https://perma.cc/3XAK-86X2>].

252. *Enforcement Query System (EQS)*, FED. ELECTION COMM'N, <http://eqs.fec.gov/eqs/searcheqs> (last visited July 22, 2017) [<https://perma.cc/96GT-FPQY>].

administrative fines for violating reporting requirements.²⁵³ Users can view the name of the party fined, the type of report filed late (or not at all), the amount of the fine, and any related candidate information.²⁵⁴

Congress has recognized that data held by the FEC can be more meaningful when combined with data held by other institutions, including Congress itself. For example, the Lobbying Disclosure Act of 1995 requires “lobbyists” to register with the Clerk of the House and the Secretary of the Senate and disclose their lobbying activities, including who they lobbied and on what issues, bills, or other government action.²⁵⁵ The Act requires House and Senate officials to use “computerized systems” with “coding” and “cross-indexing” to “maximize public access to materials filed.”²⁵⁶ Congress also requires these reports to be available over the internet.²⁵⁷ In 2007, frustrated with the slow rate of publication online, Congress amended the law to require publication online in a searchable, sortable, and downloadable format.²⁵⁸ The 2007 amendments also linked lobbying information with campaign contribution data reported to the FEC,²⁵⁹ so that users can better track financial ties between lobbyists and public officials. Thus, the amendments demonstrate how data regimes can be dynamic rather than static.

But, as with many other disclosure-based regulatory schemes, researchers question how effectively campaign finance disclosures have achieved their stated goals of preventing corruption and the appearance of it.²⁶⁰ Despite

253. *Administrative Fines*, FED. ELECTION COMM’N, <http://www.fec.gov/data/AdminFine.do?format=html> (last visited July 22, 2017) [<https://perma.cc/EB8S-VY5D>].

254. *Id.*

255. Lobbying Disclosure Act of 1995, 2 U.S.C. § 1601 (2012).

256. *Id.* § 1605.

257. *Id.* § 1605(a)(4),(9).

258. Honest Leadership and Open Government Act of 2007, Pub. L. No. 110-81, § 103(b), 121 Stat. 735, 739; *see also Lobbying Disclosure*, OFFICE OF THE CLERK, U.S. HOUSE OF REP., <http://lobbyingdisclosure.house.gov> (last visited July 22, 2017) [<https://perma.cc/2ZUH-7R8J>]; *Query the Lobbying Disclosure Act Database*, U.S. SENATE, <http://soprweb.senate.gov/index.cfm?event=selectFields&reset=1> (last visited July 22, 2017) [<https://perma.cc/YR6C-GC3P>].

259. Honest Leadership & Open Government Act, §§ 203, 204, Pub. L. No. 110-81, 121 Stat. 735, 742-46 (codified at 2 U.S.C. §§ 1604(d), 434(i)).

260. Heerwig & Shaw, *supra* note 243. Larry Lessig’s work is particularly skeptical of FEC disclosures. *See, e.g.*, LAWRENCE LESSIG, *REPUBLIC LOST: HOW MONEY CORRUPTS CONGRESS—AND A PLAN TO STOP IT* 251–60 (2011).

relatively thoughtful presentation and formatting of the data, the data themselves are not very reliable. Compliance with FEC reporting requirements is spotty, filings are often selective and incomplete (the FEC does not require filings to be complete to be accepted), FEC enforcement is limited in several important ways, data collection is not always standardized, and thus the data is often unreliable.²⁶¹ To note just one example, although the FEC tracks campaign contributions, it does not track individual contributors well because there is no unique identifier assigned to them.²⁶² Thus, the data “gives the illusion of transparency, but functions instead to obscure the most pertinent financial constituencies in a sea of data.”²⁶³ These design failures mean that the intended users—voters, intermediaries, and regulators—are relying on flawed data, even if it is presented in multiple formats.²⁶⁴

Moreover, scholars have questioned at length whether voluminous campaign finance data succeeds in achieving its stated goals and have identified a long list of preconditions necessary for the data to do so. For example, Malbin and Gais identify several requirements that align with the wisdom applied to other disclosure-based regimes—the disclosure must be accurate, usable, and accessible, both to the intermediaries who can synthesize it, and to voters who might cast votes based on it.²⁶⁵ The preconditions are substantially similar to the conditions prescribed by Fung, Graham, and Weil (which I discuss in more detail in Part V, *infra*).²⁶⁶ Numerous articles by campaign finance scholars suggest specific improvements to the FEC’s data practices, including how it collects, processes, and disseminates data, and enforces compliance.²⁶⁷ Simple fixes, like assigning unique identifiers to contributors and

261. Heerwig & Shaw, *supra* note 243, at 1479–85.

262. *Id.* at 1484.

263. *Id.*

264. *Id.* at 1486–89.

265. MICHAEL J. MALBIN & THOMAS L. GAIS, *THE DAY AFTER REFORM: SOBERING CAMPAIGN FINANCE LESSONS FROM THE AMERICAN STATES* 36 (1998).

266. Heerwig & Shaw, *supra* note 243, at 1475 (citing FUNG ET AL., *supra* note 15, at 11).

267. For just a small sampling, see, e.g., Heerwig & Shaw, *supra* note 243; Hasen, *supra* note 243; Richard Briffault, *Updating Disclosure for the New Era of Independent Spending*, 27 J.L. & POL. 683 (2012).

allowing the FEC to conduct random audits, might greatly improve the quality of the data.²⁶⁸

Finally, as with other data regimes, third-party intermediaries play an important role in translating and synthesizing the government's campaign finance data. Groups like the Center for Responsive Politics (CRP), the Campaign Finance Institute (CFI), and the Sunlight Foundation rely on FEC data to provide more digestible information to the public.²⁶⁹ Indeed, an old 2002 survey found that political journalists relied on the CRP website more than any other, with the FEC's own site ranking third.²⁷⁰ Another group called MAPlight.org mashes up the publicly available voting records of members of Congress with campaign finance data, trying to find correlations.²⁷¹ MapLight gathers data from not only the FEC, but also from the CRP, which runs OpenSecrets.org and FollowTheMoney.org (for California data).²⁷² However, because the FEC data is so flawed, scholars worry that it is not suitable for use by researchers and other informational intermediaries who might otherwise be able to extrapolate larger patterns or trends.²⁷³ Thus, the FEC's regime is characterized by noble sentiments but flawed data.

Therefore, although there is low confidence in the FEC's data sets, there remains clear demand for such data and somewhat feasible fixes that are likely to appeal to many interested parties, such as voters, watchdogs, and the disclosers themselves. Improved data quality—using the methods recommended in Part V and the recommendations of earlier projects—should enable better, more frequent uses of the information and perhaps even affect disclosers' behavior.

268. Heerwig & Shaw, *supra* note 243, at 1494–99.

269. *Id.* at 1487–88.

270. *Id.* (citing ALBERT L. MAY, INST. FOR POLITICS, DEMOCRACY, & THE INTERNET, *THE VIRTUAL TRAIL: POLITICAL JOURNALISM ON THE INTERNET* 22 (2002), http://www.pewtrusts.org/~media/legacy/uploadedfiles/wwwpewtrustsorg/reports/the_practice_of_journalism/pponlinejournalistpdf.pdf [<https://perma.cc/H3X2-M8JK>]).

271. MAPLIGHT, <http://www.maplight.org> (last visited July 22, 2017) [<https://perma.cc/5EFN-MCCA>].

272. *Find Contributions*, MAPLIGHT, <http://maplight.org/us-congress/contributions> (last visited July 22, 2017) [<https://perma.cc/7T2W-8S8Z>].

273. Heerwig & Shaw, *supra* note 243, at 1490.

B. The EPA's Toxic Release Data

Like campaign finance law, environmental law relies on public disclosure to regulate underlying conduct. Environmental scholars sometimes refer to this as “regulation by revelation”—leveraging the threat of public backlash to change the underlying behavior that leads to pollution.²⁷⁴ As the EPA’s former General Counsel observed, “[i]nformation . . . can be a supplement, sometimes even an alternative, to regulation. When broadly available, information can change behavior.”²⁷⁵

This logic undergirds several well-known environmental statutes that are predicated on disclosure, such as the Clean Air Act; the Emergency Planning and Community Right-to-Know Act; the Federal Insecticide, Fungicide, and Rodenticide Act; the National Environmental Policy Act; the Safe Drinking Water Act; and the Toxic Substances Control Act.²⁷⁶ The commonality is that these statutes all “place affirmative duties on federal agencies to make information available to the public.”²⁷⁷ In combination, these laws “seem to provide a right of public access to virtually all environmental information in the hands of the federal government.”²⁷⁸

In fact, perhaps the most well-known agency database dates back to 1986, when Congress passed the Emergency Planning and Community Right-to-Know Act, requiring the EPA to establish a national toxic chemical inventory, with the information “in a computer data base . . . accessible to any person.”²⁷⁹ The database became the Toxics Release Inventory (TRI) program.

The TRI program requires facilities to report their production and release of roughly 650 dangerous chemicals.²⁸⁰

274. Esty, *supra* note 4, at 126; ANN FLORINI, *THE COMING DEMOCRACY: NEW RULES FOR RUNNING A NEW WORLD* 188–90 (2003); William F. Pederson, *Regulation and Information Disclosure: Parallel Universes and Beyond*, 25 HARV. ENVTL. L. REV. 151, 160–61(2001).

275. Frank S. Arnold, *Reduced Health a Regulatory Cost?*, 15 ENVIRO. F. 36 (July/Aug. 1998).

276. Vladeck, *supra* note 64, at 1788 (citations omitted).

277. *Id.*

278. *Id.*

279. Emergency Planning and Community Right-to-Know Act, Pub. L. No. 99-499, § 313 (codified as amended at 42 U.S.C. §§ 11023(h), (j)).

280. 30 C.F.R. § 372.65 (2016); *see also* 42 U.S.C. § 11023.

The EPA first reported TRI data in 1989, and first published it online in 1998.²⁸¹ Today's version allows users to search for toxic release data by state, county, city, or ZIP code, and will generate a customized "factsheet" based on the query, listing all facilities in the geographic area and the quantity of chemicals they release.²⁸² The data are presented in colorful chart, graph, and map forms.²⁸³ For each reporting facility, the EPA maintains a "Facility Profile Report" with more granular data regarding the amount of chemicals managed, released, or transferred.²⁸⁴ Still more data is available by link to each company's full reports in the EPA's Envirofacts database.²⁸⁵

TRI has been widely hailed for having a "significant impact on firm-level emissions" and even inspiring several other disclosure-based regulatory efforts, both in the United States and overseas.²⁸⁶ The initial success of the TRI program even surprised the EPA and environmental groups who had toiled for years to regulate toxic pollution.²⁸⁷ Ten years after initiating the TRI program, the amount of pollution released had dropped by half.²⁸⁸ In fact, initial media interest in the program and the resulting threat of negative publicity appeared to have a powerful impact on companies—even before the first reports were required, executives of some companies promised to reduce their toxic outputs by as much as 90 percent.²⁸⁹ As one of the earliest programs of its kind, the TRI

281. *TRI Program Milestones*, ENVTL. PROT. AGENCY, http://www.epa.gov/sites/production/files/documents/tritimeline_0.pdf [https://perma.cc/VZJ9-849B].

282. *Toxics Release Inventory (TRI) Program*, ENVTL. PROT. AGENCY, <http://www.epa.gov/toxics-release-inventory-tri-program> (last visited July 22, 2017) [https://perma.cc/Y23G-LD3T].

283. *See, e.g., 2014 TRI Factsheet: City – Dallas, TX*, ENVTL. PROT. AGENCY, https://iaspub.epa.gov/triexplorer/tri_factsheet.factsheet?&pstate=TX&pcity=Dallas&pyear=2014&pDataSet=TRIQ1 (last visited July 22, 2017) [https://perma.cc/9HRC-WBPK].

284. *See, e.g., Facility Profile Report*, ENVTL. PROT. AGENCY, https://iaspub.epa.gov/triexplorer/release_fac_profile?TRILIB=TRIQ1&FLD=&FLD=RELLBY&FLD=TSFDSP&OFFDISPD=&OTHDISPD=&ONDISPD=&OTHOFFD=&tri=75209TXSNS6000L&YEAR=2014 (last visited Aug. 6, 2017) [https://perma.cc/YH97-5AJE].

285. *Envirofacts*, ENVTL. PROT. AGENCY, <https://www3.epa.gov/enviro/index.html> (last visited July 22, 2017) [https://perma.cc/GF6N-8H6C].

286. Mark A. Cohen & W. Kip Viscusi, *The Role of Information Disclosure in Climate Mitigation Policy*, 3 CLIMATE CHANGE ECO. 1250020-1-2 (2012).

287. FUNG ET AL., *supra* note 15, at 29.

288. *Id.*

289. *Id.*

has been applauded by many as the best example of regulation via disclosure.²⁹⁰

Perhaps inspired by its own success, the EPA now publishes hundreds of datasets online. The EPA website publishes so many datasets that it includes several landing pages that help users search for and navigate the data available.²⁹¹ Some of the more well-known datasets after TRI include the EPA's Enforcement and Compliance History Online (ECHO) website, which allows users to search for recent and historical enforcement actions, including the last date of inspection.²⁹² The EPA currently lists ninety-six datasets on its website,²⁹³ with 1,738 listed on Data.gov.²⁹⁴ The agency even hosts an online discussion forum for data developers.²⁹⁵

Despite the initial success of TRI, Fung, Graham, and Weil found that in comparison with seven other disclosure regimes, toxic release disclosure was only moderately successful, at best.²⁹⁶ They found, in particular, that toxic release data is not embedded in potential users' decision-making, as "[m]ost home buyers, renters, job seekers, consumers, and investors do not consider toxic pollution" when making decisions.²⁹⁷ The TRI data did succeed in better informing policymakers, such as Congress and the EPA itself.²⁹⁸ And many manufacturers quickly embedded the new data into their decision-making, recognizing the reputational and regulatory consequences.²⁹⁹

290. See, e.g., Esty, *supra* note 4; Pederson, *supra* note 274.

291. See, e.g., *Environmental Dataset Gateway (EDG)*, ENVTL. PROT. AGENCY, <https://edg.epa.gov/metadata/catalog/main/home.page> (last visited July 22, 2017) [<https://perma.cc/5K28-9YM3>]; *Envirofacts*, *supra* note 285; Env'tl. Prot. Agency, *Data Finder*, EPA.GOV, <https://iaspub.epa.gov/enviro/datafinder.agency> (last visited July 22, 2017) [<https://perma.cc/XPU4-7MF9>].

292. *Enforcement and Compliance History Online (ECHO)*, ENVTL. PROT. AGENCY, <https://echo.epa.gov> (last visited July 22, 2017) [<https://perma.cc/9YZ6-85CX>].

293. *Developer Central: Data*, ENVTL. PROT. AGENCY, <https://developer.epa.gov/category/data/> (last visited July 22, 2017) [<https://perma.cc/4HTQ-5KJ8>].

294. *EPA Data Catalog*, DATA.GOV, http://catalog.data.gov/dataset?organization=epa-gov&organization_limit=0 (last visited on Mar. 11, 2016) [<https://perma.cc/6KJD-USKE>].

295. *Data and Developer Forum*, ENVTL. PROT. AGENCY, <http://blog.epa.gov/data/> (last visited July 22, 2017) [<https://perma.cc/UK3C-WJPQ>]. The forum does not appear to be particularly active.

296. FUNG ET AL., *supra* note 15, at 76, 79.

297. *Id.* at 85.

298. *Id.* at 86.

299. *Id.*

But researchers have gradually curbed their enthusiasm based on studies showing flaws in reporting (including inaccurate data) and longitudinal studies showing less impact on potential users' and disclosers' actual conduct.³⁰⁰ Scholars have long worried that EPA datasets are "patchy" and "unreliable."³⁰¹

In the last decade, despite its success, the EPA has "drastically scaled back the information made public" under the TRI program.³⁰² Prior to 2006, the EPA required facilities to report information regarding any chemical release over 500 pounds.³⁰³ But a 2006 rule increased the threshold to 5,000 pounds, provided the total annual release into the environment does not exceed 2,000 pounds.³⁰⁴ The GAO criticized the EPA's rule as reducing the "quantity and detail of information" released to communities.³⁰⁵ Under the Trump administration, one can envision even more drastic changes to the TRI program.

But these stories have not deterred scholars and policymakers who still believe that data is the path to environmental regulation. For example, Daniel Esty argues that "[a]s data become easier to analyze and disseminate, and dramatically less costly to acquire and use, our capacity to identify and solve environmental problems will increase substantially."³⁰⁶ Indeed, he predicts that information technologies will enable an "environmental revolution perhaps as important as that which launched the modern environmental movement" in the 1960s.³⁰⁷

Of course, even data optimists believe that there are significant challenges in producing environmental data that are reliable and usable.³⁰⁸ But, as in other fields, environmental scholars see great promise in third-party watchdogs and data intermediaries translating voluminous

300. *Id.* (citing studies).

301. Esty, *supra* note 4, at 156.

302. Vladeck, *supra* note 64, at 1791.

303. Toxics Release Inventory Burden Reduction Final Rule, 71 Fed. Reg. 76,932, 76,933 (Dec. 22, 2006) (codified at 40 C.F.R. pt. 372).

304. *Id.* at 76,937.

305. JOHN B. STEPHENSON, U.S. GOV'T ACCOUNTABILITY OFFICE, ENVIRONMENTAL RIGHT-TO-KNOW: EPA'S RECENT RULE COULD REDUCE AVAILABILITY OF TOXIC CHEMICAL INFORMATION USED TO ASSESS ENVIRONMENTAL JUSTICE 16 (2007); Vladeck, *supra* note 64, at 1791 n.33.

306. Esty, *supra* note 4, at 119.

307. *Id.*

308. *Id.* at 171-74.

data into usable heuristics for consumers, perhaps creating a race to the top among regulated firms.³⁰⁹ Nevertheless, not even the most well-resourced and well-meaning intermediaries can cure flawed data. The EPA's many data sets, then, suffer more for quality than for quantity.

C. *The CFPB's Consumer Complaint Data*

The Consumer Financial Protection Bureau (CFPB) is the newest agency among those surveyed here, and its newness makes it interesting. Born of the 2008 financial crisis, the Bureau was "designed in a world of new technology."³¹⁰ In short, the CFPB "is a new agency operating under a new statute and is on the frontier of the open data trend."³¹¹ The Bureau's Office of Consumer Response operates a process for consumers to file complaints regarding financial products and services, which quite notably are published online by the CFPB in a massive, searchable Consumer Complaint Database, identifying companies by name.³¹²

The CFPB website allows consumers to file complaints for eleven categories of financial products, including mortgages, student loans, and credit cards.³¹³ Complaints can specify the name of the company, the type of product or service at issue, the type of problem with it, and the consumer's ZIP code, all of which the Bureau authenticates and sends to the company for a response.³¹⁴ The published data is searchable and sortable,

309. *Id.* at 208–09.

310. Katherine Porter, *The Complaint Conundrum: Thoughts on the CFPB's Complaint Mechanism*, 7 BROOK. J. CORP. FIN. & COM. L. 57, 86 (2012).

311. CORTEZ, *supra* note 143, at 62 (quotation marks omitted). The Bureau was created by Title X of the Dodd-Frank Wall Street Reform and Consumer Protection Act of 2010, Pub. L. No. 111-203, 124 Stat. 1376 (2010) (codified in scattered sections of 12 U.S.C.). The idea for the CFPB germinated, of course, from a 2007 essay by then-professor Elizabeth Warren. Elizabeth Warren, *Unsafe at Any Rate*, DEMOCRACY (Summer 2007), <https://democracyjournal.org/magazine/5/unsafe-at-any-rate/> [<https://perma.cc/SX8N-6G3X>].

312. *Consumer Complaint Database*, CONSUMER FIN. PROT. BUREAU, <http://www.consumerfinance.gov/complaintdatabase/> (last visited July 22, 2017) [<https://perma.cc/Q48G-ZRHZ>] [hereinafter *Consumer Complaint Database*].

313. *Id.* The database initially posted complaints only for credit cards. *See* Disclosure of Certain Credit Card Complaint Data, 76 Fed. Reg. 76,628 (Dec. 8, 2011). In 2012, the Bureau broadened the database to include other products and services regulated by the Bureau. Disclosure of Consumer Complaint Data, 77 Fed. Reg. 37,616 (June 22, 2012).

314. Disclosure of Consumer Complaint Data, 77 Fed. Reg. at 37,616–17.

and includes not only the company's name, but also its response (if any) and whether the response was timely or further disputed by the customer.³¹⁵ Companies select their responses via a pull down menu that includes options such as "Closed with monetary relief," "Closed with non-monetary relief," "Incorrect company," and "In progress."³¹⁶ Companies have a total of sixty days to respond, and late responses are tagged by the CFPB as "Past due" or "No response" if the delay exceeds thirty days.³¹⁷ Each complaint and response is published, but only if it meets numerous publication criteria.³¹⁸

In 2014, the Bureau expanded the database (reversing its previous position) by posting narrative commentary by consumers.³¹⁹ It originally declined to do so, citing privacy and the risk of disclosing consumers' personal information.³²⁰ But after considering industry objections (including the potential harm to company reputations) and devising ways to scrub the information of personally identifiable information, the Bureau finalized its plan to include consumer narratives in the Complaint Database.³²¹ As a measure of symmetry, the Bureau proposed to allow companies to post their own narrative responses,³²² but companies preferred to respond with a preset list of "structured" responses, such as "Company acted appropriately," "Factual dispute," "Misunderstanding," and "Opportunity for improvement."³²³ However, these responses

315. CORTEZ, *supra* note 143, at 62–63.

316. CONSUMER FIN. PROT. BUREAU, COMPANY PORTAL MANUAL (V. 2.14) 2–3 (2015), <http://www.cfbblaw.com/files/Uploads/Documents/Resources/cfpb-company-portal-manual-may-2015.pdf> [<https://perma.cc/3MRX-TYBR>] [hereinafter COMPANY PORTAL MANUAL].

317. *Id.* at 24.

318. The Bureau will not publish complaints if they are missing critical information, have been referred to other regulators, are duplicative, would reveal trade secrets, are fraudulent, or identify the wrong company. *Id.* at 26; Final Policy Statement: Disclosure of Consumer Complaint Data, 78 Fed. Reg. 21,218, 21,225 (Apr. 10, 2013).

319. See Disclosure of Consumer Complaint Narrative Data, Notice of Proposed Policy Statement with Request for Public Comment, 79 Fed. Reg. 42,765 (July 23, 2014).

320. Disclosure of Consumer Complaint Data, 77 Fed. Reg. at 37,568.

321. See Final Policy Statement: Disclosure of Consumer Complaint Narrative Data, 80 Fed. Reg. 15,572 (Mar. 24, 2015).

322. Disclosure of Consumer Complaint Narrative Data, Notice of Proposed Policy Statement with Request for Public Comment, 79 Fed. Reg. at 42,768.

323. COMPANY PORTAL MANUAL, *supra* note 316, at 29.

are optional; companies need not select one for publication.³²⁴ Today, then, the Complaint Database includes narrative descriptions of consumers' problems (if they choose to narrate them), which can make the problems more concrete and compelling than displaying relatively sanitized data entries alone.³²⁵

The intended users of the Complaint Database are consumers, researchers, the Bureau, other regulators, and even the subjects of the complaints themselves—companies.³²⁶ The Bureau and consumer groups emphasize that the primary intended beneficiaries are consumers, and publishing complaint data is a “public service” that can “empower” consumers and help them avoid “bad actors” in these markets.³²⁷ The Bureau itself emphasizes that “disclosure is one of the best tools government agencies can use.”³²⁸ Former Bureau Director Richard Cordray encouraged “the public, including consumers, the companies that serve them, analysts, data scientists, civic hackers, developers, policymakers, journalists, and academics, to analyze, augment, and build on the public database.”³²⁹ Bureau staff also hope that intermediaries develop mobile apps and other information products based on complaint data.³³⁰

This latter aspiration—that intermediaries will use the government's data—is being realized, at least modestly so far. Academics are publishing empirical analyses of the CFPB's data.³³¹ Public interest research groups (PIRGs) are producing reports of certain financial product categories, like credit cards

324. Final Policy Statement: Disclosure of Consumer Complaint Narrative Data, 80 Fed. Reg. at 15,583.

325. CORTEZ, *supra* note 143, at 65.

326. *Id.*; Disclosure of Certain Credit Card Complaint Data, 76 Fed. Reg. 76,628, 76,630–31 (Dec. 8, 2011).

327. Final Policy Statement: Disclosure of Consumer Complaint Data, 78 Fed. Reg. 21,220 (Apr. 10, 2013).

328. *Id.* at 21,220.

329. Richard Cordray, Director, Consumer Fin. Prot. Bureau, Prepared Remarks of Director Richard Cordray at the Consumer Response Field (Mar. 28, 2013), <http://www.consumerfinance.gov/speeches/prepared-remarks-of-director-richard-cordray-at-the-consumer-response-field-hearing/> [<https://perma.cc/J8LF-5S7R>].

330. CORTEZ, *supra* note 143, at 66.

331. See, e.g., Ian Ayres et al., *Skeletons in the Database: An Early Analysis of the CFPB's Consumer Complaints*, 19 FORDHAM J. CORP. & FIN. L. 343 (2014).

and debt collection, based on Bureau data.³³² In 2015, the rankings-crazed *U.S. News & World Report* ranked credit cards by relying, in part, on data from the Consumer Complaint Database.³³³

Similar to other agencies, the CFPB hopes that the act of data publication itself will encourage companies to improve their underlying behavior.³³⁴ Bureau staff report that complaints have inspired some companies to address potential problems of their own, such as long customer service phone trees.³³⁵ Indeed, one of the Bureau's original aspirations was that published complaints would encourage companies to compete in a race to the top, based on how they handled customer service and customer complaints.³³⁶ Management consulting firms now advise companies to "turn what they hear from the CFPB's consumer complaint database into a business advantage."³³⁷ Bureau staff also report that some companies have tied executive compensation to how well the company has responded to published complaints.³³⁸ In short, the Bureau sees its role somewhat modestly as a publisher—providing a window into a dialogue between companies and their customers.³³⁹ Once complaints are published, "[t]he marketplace of ideas then does the rest."³⁴⁰

Of course, the Consumer Complaint Database has not been without controversy, particularly to the firms identified in it. Firms and industry groups filed scores of public comments objecting to Bureau proposals to publish complaint data. The objections fall into seven general categories: (i) it is unfair to

332. See, e.g., *Reports: The CFPB Gets Results for Consumers*, U.S. PUB. INT. RESEARCH GRP. (PIRG), [http://www.uspirg.org/page/usf/reports-cfpb-gets-result s-consumers](http://www.uspirg.org/page/usf/reports-cfpb-gets-result-s-consumers) (last visited July 22, 2017) [<https://perma.cc/4L5L-EHT6>].

333. U.S. NEWS & WORLD REP., *U.S. News & World Report Releases the Best Credit Card Rankings*, (Nov. 10, 2015, 11:21 AM), <https://www.usnews.com/info/blogs/press-room/2015/11/10/us-news-releases-the-best-credit-card-rankings> [<https://perma.cc/2Q2E-W5ZP>].

334. CORTEZ, *supra* note 143, at 66.

335. *Id.*

336. Disclosure of Certain Credit Card Complaint Data, 76 Fed. Reg. 76,628, 76,630 (Dec. 8, 2011).

337. *Analysis: CFPB's Consumer Complaint Database: Deloitte's Analysis Reveals Valuable Insights*, DELOITTE, <http://www2.deloitte.com/us/en/pages/financial-services/articles/consumer-financial-protection-bureau-cfpb-consumer-complaint-database.html> (last visited July 22, 2017) [<https://perma.cc/63V7-S752>].

338. CORTEZ, *supra* note 143, at 66.

339. *Id.* at 66–67.

340. Disclosure of Certain Credit Card Complaint Data, 76 Fed. Reg. at 76,631.

publish complaints that are not verified by the Bureau; (ii) the complaints are self-selecting and thus are non-random and non-representative of customer experiences; (iii) the data lack context and might appear to be endorsed by the Bureau; (iv) the data are susceptible to manipulation and fraud; (v) the companies will suffer reputational harm in the media and might draw the attention of plaintiffs' lawyers; (vi) the database is overinclusive because it includes complaints that are not necessarily legal or regulatory violations; and (vii) the Bureau lacks statutory authority to publish complaint data online.³⁴¹ The Bureau responded at length to these objections in the Federal Register, showing a basic sensitivity to industry concerns, though disagreeing with industry conclusions that the Bureau should not publish the data online.³⁴²

Moreover, the Bureau has fielded industry complaints about the database in several formats, including in public comments filed during notice and comment periods, in letters to the Director, and in complaints to the Bureau's Ombudsman.³⁴³ The Federal Reserve's Office of Inspector General, which has oversight responsibility for the CFPB, has audited the database "to assess the effectiveness of the [CFPB's] controls over the accuracy and completeness of its public-facing Consumer Complaint Database."³⁴⁴ Still, the database remains a target for industry, and a bill proposed in the 115th Congress would remove it from the public domain,³⁴⁵ sparking an outcry from consumer watchdogs and public interest groups.³⁴⁶

Thus, as a new database being run by a new agency, the Consumer Complaint Database continues to be refined and new uses (and objections) continue to emerge. Although the database seems well-designed to defuse potential criticisms, it

341. CORTEZ, *supra* note 143, at 67–70.

342. *Id.* at 67–71.

343. *Id.* at 75.

344. FED. RESERVE OFFICE OF INSPECTOR GEN., 2015-FMIC-C-016, OPPORTUNITIES EXIST TO ENHANCE MANAGEMENT CONTROLS OVER THE CFPB'S CONSUMER COMPLAINT DATABASE, AUDIT REPORT 1 (2015).

345. Financial CHOICE Act of 2017, H.R. 10, 115th Cong. § 725 (2017–18). H.R. 10 passed the House on June 8, 2017.

346. See, e.g., Ed Mierzwinski, *Financial Choice Act: A Cruel Choice for the CFPB & Consumers*, U.S. PIRG (Apr. 24, 2017), <http://www.uspirg.org/blogs/eds-blog/usp/financial-choice-act-cruel-choice-cfpb-consumers> [<https://perma.cc/8J75-GX73>].

is still susceptible to being undermined by an unsympathetic Trump administration and a Republican-led Congress, per the pending bill.³⁴⁷ Nevertheless, the CFPB's efforts can serve as a model for other agencies considering publishing consumer reports or complaints.³⁴⁸

D. The CPSC's Product Safety Data

The Consumer Product Safety Commission's database, SaferProducts.gov, enjoys more clear statutory authority and attention from Congress than most agency databases. Since the 1970s, Congress has required the CPSC to "protect the public against unreasonable risks of injury" and "assist consumers in evaluating the comparative safety of consumer products."³⁴⁹ But in 2008, Congress passed the Consumer Product Safety Improvement Act, requiring the Commission to create a searchable online database of product safety incidents.³⁵⁰ The law required the database to include "reports of harm relating to the use of consumer products," including reports from consumers, physicians, state and local governments, and others.³⁵¹ Each report must describe the product or substance at issue, identify the manufacturer or labeler, and describe the harm reported.³⁵² Supporters hailed the effort to "empower consumers," "expedite recall disclosure," and "enhance a family's right to know about dangerous and defective products on the market."³⁵³ The CPSC launched SaferProducts.gov in March 2011.³⁵⁴

Today, SaferProducts.gov includes a searchable online database of thousands of "reports of harm" related to identified

347. H.R. 10.

348. See, e.g., ACUS, Adoption of Recommendations, 81 Fed. Reg. 40,259 (June 21, 2016).

349. 15 U.S.C. § 2051(b) (2012).

350. Consumer Product Safety Improvement Act of 2008 (CPSIA), Pub. L. No. 110-314, § 212, 122 Stat. 3016 (2008) (codified at 15 U.S.C. § 2055a).

351. 15 U.S.C. § 2055a(b)(1)(A) (2012).

352. 15 U.S.C. § 2055a(b)(2)(B). U.S. Consumer Product Safety Commission (CPSC) regulations provide more detail. 16 C.F.R. § 1102.10(d) (2011).

353. 154 CONG. REC. 7577-01 (2008) (statement by Rep. Eshoo); 1645-01 (2008) (statement by Rep. Delauro); 7577-01 (2008) (statement by Rep. Hollen).

354. CPSC published the final rule in 2010. Publicly Available Consumer Product Safety Information Database, 75 Fed. Reg. 76,832 (Dec. 9, 2010) (codified at 16 C.F.R. pt. 1102).

products.³⁵⁵ Like the CFPB's Consumer Complaint Database, the site has a portal for users to report incidents and a portal for companies to respond to them.³⁵⁶ The database is keyword-searchable with advanced search options that include product name, company or brand name, and the product model.³⁵⁷ The advanced search also allows users to search for incidents by date, location, the "Victim's Age," and by "Injury Information," including reports of death.³⁵⁸

Interestingly, the Act also requires the CPSC to "provide clear and conspicuous notice to users of the database that the Commission does not guarantee the accuracy, completeness, or adequacy of the contents of the database."³⁵⁹ As such, SaferProducts.gov includes a disclaimer that tracks this language almost verbatim.³⁶⁰

An innovation in SaferProducts.gov that might be emulated by other agency databases³⁶¹ is that the CPSC allows manufacturers to comment on reports and object to inaccuracies. This feature derives from the Act itself, which dictates that the CPSC "shall" provide manufacturers and labelers an opportunity to comment on incident reports and request that such comments be included in the report posted

355. An accounting on July 25, 2013, found 15,517 reports on SaferProducts.gov (roughly 18 months after the site went live on March 11, 2011). See *The Consumer Product Incident Database — Saferproducts.gov*, COALITION FOR SENSIBLE SAFEGUARDS (Aug. 5, 2013), <http://www.sensible safeguards.org/fact-sheets/the-consumer-product-incident-database/> [<https://perma.cc/V2UP-Q4WJ>].

356. The site includes "portals" for both consumers and businesses. See, e.g., U.S. Consumer Prod. Safety Comm'n, *File a Report*, SAFERPRODUCTS.GOV, <https://www.saferproducts.gov/CPSRMSPublic/Incidents/ReportIncident.aspx> (last visited Aug. 6, 2017) [<https://perma.cc/Z4ZE-RJBZ>]; U.S. Consumer Prod. Safety Comm'n, *Business Portal*, SAFERPRODUCTS.GOV, <https://www.saferproducts.gov/CPSRMSPublic/Industry/Home.aspx> (last visited Aug. 6, 2017) [<https://perma.cc/T3T7-XFZE>].

357. U.S. Consumer Prod. Safety Comm'n, *Advanced Search*, SAFERPRODUCTS.GOV, <http://www.saferproducts.gov/Search/AdvancedSearch.aspx> (last visited July 16, 2017) [<https://perma.cc/2Q2K-V379>].

358. *Id.*

359. 15 U.S.C. § 2055a(b)(5) (2012).

360. U.S. Consumer Prod. Safety Comm'n, *Home Page*, SAFERPRODUCTS.GOV, <http://www.saferproducts.gov/Default.aspx> (last visited July 16, 2017) [<https://perma.cc/Y784-6L4N>] ("CPSC does not guarantee the accuracy, completeness, or adequacy of the contents of the Publicly Available Consumer Product Safety Information Database on SaferProducts.gov, particularly with respect to information submitted by people outside of CPSC.")

361. CORTEZ, *supra* note 143, at 20–21.

online.³⁶² Likewise, the CPSC must consider objections that any information in a report is “materially inaccurate,”³⁶³ which the CPSC defines as information “that is false or misleading, and which is so substantial and important as to affect a reasonable consumer’s decision-making about the product.”³⁶⁴ Congressional attention to the validity of CPSC reports is not an accident. A series of inaccurate product safety warnings inspired Congress to amend the Consumer Product Safety Act in 1981 to improve procedural safeguards for Commission announcements.³⁶⁵

The procedures for SaferProducts.gov are already being tested. In a recent case, *Company Doe v. Tenenbaum*, a company anonymously challenged an inaccurate product safety report in the database.³⁶⁶ The court found that the CPSC database “bears the Government’s stamp of approval through its publication on an official website that, by its terms, is a repository of reports regarding ‘unsafe products.’”³⁶⁷ The court sustained the company’s challenge, somewhat remarkably,³⁶⁸ finding that the CPSC posting “materially inaccurate” information on SaferProducts.gov was not only “final agency action” under the Administrative Procedure Act (APA), but also violated the CPSC’s own regulations.³⁶⁹ In the vast majority of similar cases, courts find that negative statements or other adverse public disclosures by agencies are not “final” under the APA and thus cannot support a cause of action against the government.³⁷⁰ The CPSC litigation thus could provide a roadmap for other litigants similarly aggrieved.

Although SaferProducts.gov’s pre- and post-publication procedures might serve as a useful model for other agency databases, challenges certainly remain. For instance, it is not

362. 15 U.S.C. § 2055a(c)(2).

363. *Id.* § 2055a(c)(4).

364. 16 C.F.R. § 1102.26(a)(1) (2011).

365. Omnibus Budget Reconciliation Act of 1981, Pub. L. No. 97-35, 95 Stat. 703 (1981); CORTEZ, *supra* note 143, at 11; James T. O’Reilly, *Libels on Government Websites: Exploring Remedies for Federal Internet Defamation*, 55 ADMIN. L. REV. 507, 542–43 (2003).

366. *Co. Doe v. Tenenbaum*, 900 F. Supp. 2d 572 (D. Md. 2012), *rev’d sub nom. Co. Doe v. Public Citizen*, 749 F.3d 246 (4th Cir. 2014).

367. *Id.* at 597.

368. CORTEZ, *supra* note 143, at app. C (surveying similar cases and finding few successful challenges).

369. *Tenenbaum*, 900 F. Supp. 2d at 597.

370. Cortez, *supra* note 75, at 1443–44.

clear how often consumers use the data to make purchasing decisions (perhaps the data is more useful to distributors and others in the supply chain), nor is it clear whether publishing the data inspires manufacturers to improve product safety apart from other requirements. And presentation remains critical. A GAO review found that some users were confused about the purpose of SaferProducts.gov, viewing it as a site featuring safe rather than unsafe products.³⁷¹ The “upbeat” name of the database might have contributed to such misperceptions.³⁷² The GAO also found that although the CPSC had used various methods to inform consumers about SaferProducts.gov (which at the time generated at least 100,000 page visits a month), including use of social media, the agency could do more to publicize the resource.³⁷³

Perhaps third-party informational products could leverage the data more effectively—for example, by creating mobile phone apps that can scan products in the aisle and display a brief product safety profile generated by data from SaferProducts.gov. Or, large online retailers like Amazon.com might find novel uses for such data, triangulating it with user reviews. Like other agencies, the CPSC’s open government website allows users to download raw data files, with voluminous links to the agency’s open data plans.³⁷⁴ Third-party intermediaries might be particularly helpful at translating “big” product safety data into comparative ratings accessible to ordinary consumers.

E. Medicare’s Quality Data

The Centers for Medicare and Medicaid Services (CMS) administers Medicare, our federal health insurance program for the elderly and chronically disabled. Because Medicare pays

371. U.S. GOV’T ACCOUNTABILITY OFFICE, CONSUMER PROD. SAFETY COMM’N, GAO-13-306, AWARENESS, USE, AND USEFULNESS OF SAFERPRODUCTS.GOV (2013), <http://www.gao.gov/assets/660/652916.pdf> [<https://perma.cc/Y3M6-WXMG>] [hereinafter CPSC: AWARENESS, USE, AND USEFULNESS].

372. *Id.* at 27; *Upbeat Name of SaferProducts.gov Confuses Users*, CONSUMER REPORTS (Mar. 12, 2013, 4:38 PM), <http://www.consumerreports.org/cro/news/2013/03/upbeat-name-of-saferproducts-gov-confuses-users/index.htm> [<https://perma.cc/YC57-KQYT>] (discussing the 2013 GAO report).

373. CPSC: AWARENESS, USE, AND USEFULNESS, *supra* note 371, at 12–14.

374. *The CPSC Open Government Plan*, CONSUMER PROD. SAFETY COMM’N, <http://www.cpsc.gov/> (last visited July 15, 2017) [<https://perma.cc/7SDY-6YU9>].

thousands of nongovernmental physicians, hospitals, and other types of providers for care, beneficiaries often must choose among multiple local providers—sometimes scores or even hundreds of such providers. But for a Medicare beneficiary that needs cardiac surgery, for example, choosing a specific surgeon in a specific hospital can be daunting.³⁷⁵

To facilitate such decisions, CMS operates five searchable databases that compare Medicare providers—Hospital Compare,³⁷⁶ Physician Compare,³⁷⁷ Nursing Home Compare,³⁷⁸ Home Health Compare,³⁷⁹ and Dialysis Facility Compare.³⁸⁰ The search functions all work in roughly the same way. Users can search for providers by city, state, or ZIP code, and then view a list of results within a twenty-five mile radius, each sortable by different criteria. For example, the Nursing Home and Dialysis Facility databases allow users to sort facilities based on how they rate on a five-star scale.³⁸¹ Hospitals are sortable by distance, by whether they offer emergency services, and by “hospital type.”³⁸² Home Health facilities are listed by the type of care offered, such as physical therapy and occupational therapy.³⁸³ Physician searches are more complicated (and are probably the least useful of the five),

375. For in-depth critiques of the consumer-driven health policy and the heroic assumptions it often makes about patients' capacity to understand and make important medical and spending decisions, see TIMOTHY STOLTZFUS JOST, *HEALTH CARE AT RISK: A CRITIQUE OF THE CONSUMER-DRIVEN MOVEMENT* (2007) and Mark A. Hall & Carl E. Schneider, *Patients as Consumers: Courts, Contracts, and the New Medical Marketplace*, 106 MICH. L. REV. 643 (2008).

376. *Hospital Compare*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/hospitalcompare/search.html> (last visited Aug. 6, 2017) [<https://perma.cc/SG2Q-RVBY>].

377. *Physician Compare*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/physiciancompare/search.html> (last visited Aug. 6, 2017) [<https://perma.cc/E7DN-6P59>].

378. *Nursing Home Compare*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/nursinghomecompare/search.html> (last visited Aug. 6, 2017) [<https://perma.cc/E7DN-6P59>].

379. *Home Health Compare*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/homehealthcompare/search.html> (last visited Aug. 6, 2017) [<https://perma.cc/N6Q7-K3BQ>].

380. *Dialysis Facility Compare*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/dialysisfacilitycompare/#search> (last visited Aug. 6, 2017) [<https://perma.cc/R278-XGKV>].

381. See *Nursing Home Compare*, *supra* note 378; *Dialysis Facility Compare*, *supra* note 380.

382. See *Hospital Compare*, *supra* note 376.

383. See *Home Health Compare*, *supra* note 379.

requiring the user to also search for a physician's name, specialty, or medical condition to help narrow the results.³⁸⁴

The five databases offer quite distinct data on "quality." Hospital Compare includes data on over 4,000 hospitals nationwide,³⁸⁵ allowing users to compare up to three hospitals at a time, using six categories: "Survey of Patients' Experiences," "Timely & Effective Care," "Complications, Readmissions & Deaths," "Use of Medical Imaging," and "Payment & Value of Care." Each tab, moreover, includes several subcategories of information. For example, "Timely & Effective Care" is divided into ten subcategories, such as "Heart Attack Care" and "Stroke Care."³⁸⁶ Results are compared to state and national averages as reference points.

The data displayed, however, are less satisfying than the categories might suggest. A frustrating proportion of data for hospitals is listed as "Not Available," with numbered footnotes explaining why.³⁸⁷ The tab titled "Payment & Value of Care" might tantalize health policy wonks, but unfortunately, comparative data often is not displayed directly (*e.g.*, "Get Results for This Hospital" is displayed when searching for Medicare spending per beneficiary) or meaningfully (*e.g.*, "No Different than the National Average Payment" is frequently displayed). Also, "Value of Care" metrics tend to display charts full of "No Different than the National Rate" and "No Different than the National Average Payment."

An afternoon on Hospital Compare leaves one with the impression that there are probably too many categories and subcategories of data for the average person to make meaningful comparisons, particularly when much of the data is "Not Available." Thus, although CMS touts Hospital Compare as "an important tool for individuals to use in making decisions

384. *Physician Compare: About the Data*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/physiciancompare/staticpages/data/aboutthedata.html> (last visited Aug. 6, 2017) [<https://perma.cc/6NDV-4TG7>].

385. *What is Hospital Compare?*, MEDICARE.GOV, <http://www.medicare.gov/hospitalcompare/About/What-Is-HOS.html> (last visited Aug. 6, 2017) [<https://perma.cc/ZFG3-DC4C>].

386. The ten categories are Heart Attack Care, Heart Failure Care, Pneumonia Care, Surgical Care, Emergency Department Care, Preventative Care, Children's Asthma Care, Stroke Care, Blood Clot Prevention & Treatment, and Pregnancy & Delivery Care.

387. *Hospital Compare, Footnotes*, MEDICARE.GOV, <https://www.medicare.gov/hospitalcompare/Data/Footnotes.html> (last visited Aug. 6, 2017) [<https://perma.cc/P4VF-HZ7D>].

about health care options,” it is probably more successful at meeting CMS’s other aspiration as “a way to encourage accountability of hospitals for the care they provide to patients.”³⁸⁸ Still, given how incomplete the data is, it is questionable whether it meets even this goal.

Physician quality data is even more limited, perhaps reflecting physicians’ longstanding concerns over reputation and liability.³⁸⁹ Physician Compare includes only directory-type information, such as the name and location of the physician, hospital admitting privileges, and information about physician’s education and board certifications. But there are few quality metrics. Physician Compare does feature a Physician Quality Reporting System, but it merely asks physicians to report whether they follow certain best practices.³⁹⁰ Physicians who report to CMS are then given performance scores. Unfortunately, the system is voluntary, and only some Physician Compare profiles include their performance scores.³⁹¹ Although there are over 200 reportable quality measures, few are listed on Physician Compare.³⁹²

388. Medicare Program; Changes to the Hospital Inpatient Prospective Payment Systems and Fiscal Year 2006 Rates, 70 Fed. Reg. 47,278, 47,420 (Aug. 12, 2005).

389. For a discussion of how transparency often challenges professional reputation while privacy reinforces it, including a discussion of corporate reputation, see William M. Sage, *Reputation, Malpractice Liability, and Medical Error*, in ACCOUNTABILITY: PATIENT SAFETY AND POLICY REFORM 159 (Virginia A. Sharpe, ed. 2004).

390. For a list of examples, see *Physician Compare, Physician Quality Reporting System (PQRS)*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.medicare.gov/physiciancompare/staticpages/data/pqrs.html> (last visited Aug. 6, 2017) [<https://perma.cc/834A-9JRH>].

391. *Id.*

392. The site itself explains how the data is limited:

At this time, not all health care professionals and group practices have quality measure performance scores on their Physician Compare profile page. Some health care professionals and group practices are committed to providing high quality care, but do not have quality measures. There are many reasons why health care professionals and group practices do not have quality measures for review. There are more than 200 quality measures that can be reported to CMS through multiple reporting methods. However, only certain measures reported through some of the reporting methods are currently available. Over time, more quality measures will be added to Physician Compare and more health care professionals and group practices will have measures available.

Id.

Still, the star ratings available on databases like Nursing Home Compare and Dialysis Facility Compare are easily accessible across a number of facilities and probably do help users searching for facilities nearby. Of course, Medicare quality data has many potential uses for many potential audiences, ranging from patients looking to choose the best surgeon or hospital, to providers evaluating their own performance, to policymakers seeking to understand broader trends in care.³⁹³

These databases had humble beginnings. In 2005, CMS first published ten different quality measures for hospitals across the United States, sprouting from a partnership between CMS and the Hospital Quality Alliance.³⁹⁴ In subsequent years, CMS has continued to add data from a variety of sources, including patient experience ratings, mortality rates for certain conditions, and hospital readmission rates, among many others.³⁹⁵ Today's Compare databases are an amalgam of data from a variety of sources, gradually added like ornaments to a Christmas tree.³⁹⁶

Perhaps the richest potential source of data is Medicare claims data. CMS processes over a billion Medicare claims each year,³⁹⁷ with each claim including multiple data points, including whom Medicare is paying and for what. Medicare claims data have long been used by academics, government researchers, and providers themselves to better understand the U.S. health care system.³⁹⁸ The growth of "big data" in health

393. Madison, *supra* note 30, at 1625.

394. *Id.* at 1626; *Hospital Compare*, CTRS. FOR MEDICARE & MEDICAID SERVS., <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html> (last visited July 22, 2017) [<https://perma.cc/PY8M-5H2K>]. The lineage of quality reporting in health also stretches back to precursors, like patient-centered nursing home standards in the wake of scandals involving nursing home deaths, passed under the Omnibus Budget Reconciliation Act of 1987, Pub. L. No. 100-203.

395. Madison, *supra* note 30, at 1626 (citing *Hospital Compare*, *supra* note 394).

396. As with other agencies, the raw data files at a separate site. See DATA.MEDICARE.GOV, <http://data.medicare.gov> (last visited Aug. 6, 2017) [<https://perma.cc/9TVL-DJNM>].

397. *Medicare Fee-for-Service, Payment Accuracy*, PAYMENTACCURACY.GOV, <https://paymentaccuracy.gov/program/medicare-fee-for-service/> (last visited Aug. 21, 2017) [<https://perma.cc/985Q-LGJP>]; Madison, *supra* note 30, at 1607.

398. Madison, *supra* note 30, at 1609–10.

care, in fact, roughly parallels and depends on the growth and accessibility of Medicare claims data.³⁹⁹

Although Medicare has released various data to the public for years, CMS was long restricted from releasing physician claims data by court order.⁴⁰⁰ In 2014, a year after the order was lifted, CMS released claims data for over 880,000 providers.⁴⁰¹ The trend has been followed by states—sixteen of which have created all-payer claims databases that compile claims data from almost all payers in the state, including public and private insurers.⁴⁰² However, in 2016 the Supreme Court threw a wrench in these efforts when it interpreted federal law as preempting state laws mandating reporting from self-insured employer plans.⁴⁰³ As is common for many data regimes, agencies often take two steps forward, then one step back.

Traditionally, claims data has meant merely payment data, without regard to the quality of care being provided. In 2003, Congress amended Medicare's payment formulas to encourage hospitals to report quality data,⁴⁰⁴ and today over 1,300 hospitals report data about infection rates, mortality rates, and other quality indicators in order to boost their Medicare reimbursements,⁴⁰⁵ all of which feed into the Hospital Compare site. Medicare extended these incentives from hospitals to physicians in 2006 and, by 2017 will require it.⁴⁰⁶ Yet, realizing that not everyone is covered by Medicare,

399. *Id.* at 1610.

400. *Fla. Med. Ass'n, Inc. v. Dep't of Health, Educ., & Welfare*, 947 F. Supp. 2d 1325, 1329 (M.D. Fla. 2013) (describing the longstanding injunction).

401. Madison, *supra* note 30, at 1611; Press Release, U.S. Dep't of Health & Human Servs., Historic Release of Data Gives Consumers Unprecedented Transparency on the Medical Services Physicians Provide and How Much They Are Paid (Apr. 19, 2014), <http://www.hhs.gov/news/press/2014pres/04/20140409a.html> [<https://perma.cc/WJ8R-M93R>].

402. John D. Freedman et al., *All-Payer Claims Databases – Uses and Expanded Prospects after Gobeille*, 23 N. ENG. J. MED. 375, 2016 WL 37580714 (Dec. 8, 2016).

403. *Gobeille v. Liberty Mut. Ins. Co.*, 136 S.Ct. 936 (2016) (invalidating a Vermont law requiring self-insured employers to submit claims data on grounds of ERISA preemption).

404. Medicare Prescription Drug, Improvement, and Modernization Act of 2003, Pub. L. No. 108-173, § 501(b), 117 Stat. 2066, 2289 (amending 42 U.S.C. § 1395ww(b)(3)(B)).

405. Medicare Program, 78 Fed. Reg. 50,496, 50,678–80 (Aug. 19, 2013).

406. Madison, *supra* note 30, at 1613; CTRS. FOR MEDICARE & MEDICAID SERVS., PHYSICIAN QUALITY REPORTING SYSTEM (PQRS) OVERVIEW 2–3 (Aug. 8,

Congress required CMS to release broad swaths of Medicare claims data regarding hospital care, physician care, prescription drugs, and other goods and services to enable private entities to add Medicare data to other data, on the condition that such entities generate publicly accessible quality ratings.⁴⁰⁷ Thus, various data sources are being combined in novel ways.

Nevertheless, scholars have long questioned the utility of performance data and report cards in the health industry,⁴⁰⁸ and disclosure more generally has long been a point of interest to health law scholars.⁴⁰⁹ Patients report that they seldom rely on publicly available data or even more comprehensible report cards of physicians, hospitals, or procedures.⁴¹⁰ A 2012 survey found that only 15 percent of patients reviewed online quality rankings or reviews when choosing doctors or hospitals, with the most frequent users being the most educated middle-aged users.⁴¹¹ Patients frequently are not aware of the information, do not understand it, or do not use it.⁴¹² Indeed, of the numerous disclosure regimes analyzed by Fung, Graham, and Weil, they found patient safety disclosures to be among the least effective, due to the complexity of the information, cognitive biases, the likelihood the data would be

2016), http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/PQRS_OverviewFactSheet_2013_08_06.pdf [<https://perma.cc/L3UL-UEHW>].

407. Madison, *supra* note 30, at 1614–15; Patient Protection and Affordable Care Act § 10332, 42 U.S.C. § 1395kk(e); Availability of Medicare Data for Performance Measurement, 76 Fed. Reg. 76,542, 76,567 (Dec. 7, 2011) (codified at 42 C.F.R. pt. 401); Medicare Access and CHIP Reauthorization Act of 2015, Pub. L. No. 114-10 § 105; Expanding Uses of Medicare Data by Qualified Entities, 81 Fed. Reg. 44,456 (July 7, 2016) (codified at 42 C.F.R. pt. 401).

408. There is voluminous literature on this point. *See, e.g.*, Arnold M. Epstein, *Rolling Down the Runway: The Challenges Ahead for Quality Report Cards*, 279 JAMA 1691 (1998); Martin M. Marshall et al., *The Public Release of Performance Data: What Do We Expect to Gain? A Review of Evidence*, 283 JAMA 1866 (2000); Eric C. Schneider & Arnold M. Epstein, *Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery*, 279 JAMA 1638 (1998).

409. *See, e.g.*, Sage, *supra* note 5.

410. *Id.*

411. SUSANNAH FOX & MAEVE DUGGAN, PEW RESEARCH CTR., HEALTH ONLINE 2013, at 20–23 (2013), http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf [<https://perma.cc/88VW-SYDT>].

412. Ben-Shahar & Schneider, *supra* note 17, at 672, 711–12 (citing studies).

misinterpreted, and the risk of strategic behavior by providers.⁴¹³

As a result, CMS has tried to “shrink” the voluminous and varied data by translating them into star ratings, using a five-star scale.⁴¹⁴ Five stars represents facilities that are “much above average,” four stars are “above average,” three stars are “average,” and so on.⁴¹⁵ Users can access the underlying data on the same site, including charts comparing each facility to the state and national averages.

More problematic is that studies reveal that those being measured—hospitals, physicians, and other providers—can respond in perverse ways to protect their ratings. Providers have been known to avoid sicker or more complicated patients for fear of compromising their scores on outcomes measurements.⁴¹⁶ For example, a study of cardiac surgery report cards in Pennsylvania found that cardiac surgeons responded to the new disclosure requirement by becoming more reluctant to operate on sicker patients.⁴¹⁷ The data is decidedly mixed, however. For every finding that public reporting of mortality rates reduced the rate of mortality, there are reports of selection bias by surgeons avoiding more severe, complex cases.⁴¹⁸

Moreover, despite the prevalence of doctor and hospital ratings and report cards, it is not clear whether consumers really want them: “Most consumers do not believe clinical quality varies significantly across doctors, hence the low consumer demand for clinical quality report cards.”⁴¹⁹ One

413. FUNG ET AL., *supra* note 15, at 76–77 (analyzing disclosure requirements in Pennsylvania and New York).

414. *See, e.g., Nursing Home Compare, supra* note 378; Madison, *supra* note 30, at 1627.

415. *Five-Star Quality Rating, Nursing Home Compare, MEDICARE.GOV*, <https://www.medicare.gov/nursinghomecompare/About/HowWeCalculate.html> (last visited Aug. 6, 2017) [<https://perma.cc/GP9L-QT83>].

416. *See, e.g., David Dranove et al., Is More Information Better? The Effects of “Report Cards” on Health Care Providers*, 111 J. POL. ECON. 555 (2003); Sage, *supra* note 5, at 1793.

417. Schneider & Epstein, *supra* note 408.

418. FUNG ET AL., *supra* note 15, at 89 (citing studies).

419. HA T. TU & JOHANNA R. LAUER, CTR. FOR STUDYING HEALTH SYS. CHANGE, RESEARCH BRIEF NO. 9, WORD OF MOUTH AND PHYSICIAN REFERRALS STILL DRIVE HEALTH CARE PROVIDER CHOICE 5 (2008), <http://www.hschange.org/CONTENT/1028/1028.pdf> [<https://perma.cc/UA9D-YFD6>]; Epstein, *supra* note 408, at 1694.

study found that less than one percent of patients knew how their hospital or surgeon was rated under mandated ratings systems.⁴²⁰ Thus, ratings might be better in theory than in practice, at least for now.

CMS's ratings have also generated litigation. A nursing home in Illinois sued Department of Health and Human Services (HHS) and CMS for mistakenly calculating its star rating on Nursing Home Compare, giving the facility two stars out of five rather than four.⁴²¹ The mistaken star rating was published on the CMS website, which did not correct it for almost two years.⁴²² The nursing home argued that HHS and the Illinois Department of Public Health, which conducted the underlying inspections, had violated its procedural due process rights under the Fifth and Fourteenth Amendments.⁴²³ The district court found that although the nursing home's low star rating probably did affect its reputation, "reputational harm does not require due process protection."⁴²⁴ The court agreed with the government that although a "mistaken rating could have caused some potential patients to look elsewhere for their care," it did not rise to a property interest.⁴²⁵ To qualify, the nursing home would have to show that the reputational harm also included some sort of "change in legal status," as required by the "stigma-plus" test.⁴²⁶ But the nursing home did not present evidence that there was any such change in legal status—such as a ban on referrals to the facility, a change in licensing status or reimbursement status, or some other tangible harm.⁴²⁷ Thus, the court called the mistake (and the nearly two-year delay in fixing it) "unfortunate," but not something rising to a liberty or property interest protected by due process.⁴²⁸

420. Epstein, *supra* note 408, at 1694.

421. *Bryn Mawr Care v. Sebelius*, 898 F. Supp. 2d 1009, 1011 (N.D. Ill. 2012).

422. *Id.*

423. *Id.* at 1011–12.

424. *Id.* at 1012.

425. *Id.* at 1013.

426. *Id.* at 1014.

427. *Id.* at 1014–15.

428. *Id.* at 1018. For a survey of federal opinions between 1974–2014 in which a private party challenged an agency announcement, identifying cases that invoked due process arguments, see CORTEZ, *supra* note 143, at app. C.

Still, quality ratings and other disclosure-based regulation remains all the rage in health policy,⁴²⁹ with contemporary proposals littered with patient surveys, outcomes data, star ratings or rankings, and of course federal databases like those mentioned above that combine many of these data points.⁴³⁰ Again, scholars have long been aware of the benefits, burdens, and limitations of using disclosure as a regulatory tool in health care.⁴³¹ Yet, it is possible, if not probable, that efforts to replace the Affordable Care Act⁴³² will rely heavily on disclosure to facilitate market-based reforms.

F. *The FDA's Adverse Event Data*

The U.S. Food and Drug Administration (FDA) maintains several online databases that track problems with the products and companies it regulates. For example, the agency publishes several enforcement databases that allow users to search for FDA inspections,⁴³³ warning letters,⁴³⁴ recalls,⁴³⁵ and enforcement reports.⁴³⁶ The FDA also maintains a database of good news—agency product approvals.⁴³⁷ But perhaps the most

429. In April 2016, for example, *Health Affairs* dedicated an issue to “Patients’ and Consumers’ Use of Evidence,” with several articles focused on recent data-reporting initiatives targeted at patient use. See *Patients’ and Consumers’ Use of Evidence*, 35 HEALTH AFF. 1 (2016).

430. See, e.g., FUNG ET AL., *supra* note 15, at 160–62 (listing various disclosure systems in health care).

431. See, e.g., Sage, *supra* note 5.

432. Patient Protection and Affordable Care Act, Pub. L. No. 111-148, 124 Stat. 119 (2010) (codified as amended in scattered sections of the Internal Revenue Code and 42 U.S.C.), *amended by* Health Care and Education Reconciliation Act of 2010, Pub. L. No. 111-152, 124 Stat. 1029 (2010).

433. *Inspections Database*, U.S. FOOD & DRUG ADMIN., <http://www.fda.gov/ICECI/Inspections/ucm222557.htm> (last updated July 14, 2017) [<https://perma.cc/2STT-KK5B>].

434. *Inspections, Compliance, Enforcement, and Criminal Investigations: Warning Letters*, U.S. FOOD & DRUG ADMIN., <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/> (last updated July 11, 2017) [<https://perma.cc/4K9T-AE5S>].

435. *Recalls, Market Withdrawals, & Safety Alerts*, U.S. FOOD & DRUG ADMIN., <http://www.fda.gov/Safety/Recalls/default.htm> (last updated July 14, 2017) [<https://perma.cc/PN4A-WCBQ>].

436. *Enforcement Report*, U.S. FOOD & DRUG ADMIN., http://www.accessdata.fda.gov/scripts/ires/index.cfm#tabNav_advancedSearch (last visited July 15, 2017) [<https://perma.cc/XG38-7CAT>].

437. *Drugs@FDA: FDA Approved Drug Products*, U.S. FOOD & DRUG ADMIN., <http://www.accessdata.fda.gov/scripts/cder/drugsatfd a/index.cfm> (last visited July 15, 2017) [<https://perma.cc/DT4E-6MWB>].

well-known FDA databases are those that track adverse events associated with pharmaceuticals and medical devices.

The FDA Adverse Event Reporting System (FAERS) includes a database of medication errors and adverse drug events reported to the agency.⁴³⁸ Adverse event reporting dates back at least thirty-five years, and perhaps even longer, when the agency received reports by paper.⁴³⁹ Manufacturer reports are required by regulation,⁴⁴⁰ but reports by health care professionals and consumers are only voluntary.⁴⁴¹

Evolving from previous iterations,⁴⁴² today's database remains primarily a tool for the FDA to monitor safety problems, rather than a tool aimed for use by the general public. In fact, the FAERS "database" is not really searchable to most users. The FAERS website includes aggregate statistics, as well as links to raw data files that include individual case reports.⁴⁴³ But the raw data files are published only in quarterly increments,⁴⁴⁴ and are not amenable to simple searches, as the agency notes ("A simple search of FAERS data cannot be performed with these files by persons who are not familiar with creation of relational databases").⁴⁴⁵

438. *FDA Adverse Event Reporting System (FAERS): Latest Quarterly Data Files*, U.S. FOOD & DRUG ADMIN., <https://www.fda.gov/drugs/guidancecompliance/regulatoryinformation/surveillance/adversedrugs/effects/ucm082193.htm> (last updated June 7, 2017) [<https://perma.cc/4UC2-BGYP>].

439. Postmarketing Safety Reports for Human Drug and Biological Products; Electronic Submission Requirements, 79 Fed. Reg. 33,072, 33,073 (June 10, 2014). Indeed, one source notes that the FDA has maintained an adverse event database in some form since 1967. CORTEZ, *supra* note 143, at 44 (citing David Gortler, *Adverse Event Databases (AERS Database)*, <http://50.63.91.31/FDA-adverse-event-database.html> (last visited Aug. 6, 2017) [<https://perma.cc/6CQG-7MMP>]).

440. 21 C.F.R. §§ 310.305 (adverse event reports for drugs marketed without approved new drug approval applications), 312.32 (investigational drug safety reports), 314.80 (postmarketing reporting).

441. *Questions and Answers on FDA's Adverse Event Reporting System (FAERS)*, U.S. FOOD & DRUG ADMIN., <https://www.fda.gov/drugs/guidancecompliance/regulatoryinformation/surveillance/adversedrugs/effects/> (last updated May 5, 2016) [<https://perma.cc/2C5Q-R3SB>] [hereinafter *FAERS*].

442. In September 2012, FAERS replaced the previous Adverse Event Reporting System (AERS), which itself derives from an earlier reboot of the FDA's pharmacovigilance system. *See* 74 Fed. Reg. 42,184, 42,185 (Aug. 21, 2009); 63 Fed. Reg. 65,000, 65,030 (Nov. 24, 1998).

443. *FAERS*, *supra* note 441.

444. *FDA Adverse Event Report System (FAERS), Latest Quarterly Data Files*, U.S. FOOD & DRUG ADMIN., <https://www.fda.gov/drugs/guidancecompliance/regulatoryinformation/surveillance/adversedrugs/effects/ucm082193.htm> (last updated June 7, 2017) [<https://perma.cc/4UC2-BGYP>].

445. *FAERS*, *supra* note 441.

The FDA also instructs potential users to request individual case safety reports by submitting a FOIA request.⁴⁴⁶ Individual reports are accessible, however, by searching the FDA's MedWatch website, which aggregates "clinically important safety information" for "human medical products," including drugs, devices, and biologics.⁴⁴⁷

The device counterpart to FAERS is MAUDE (Manufacturer and User Facility Device Experience), a database of device adverse events,⁴⁴⁸ which the agency began collecting in 1984.⁴⁴⁹ Today, MAUDE includes both mandatory and voluntary adverse event reports, with downloadable data files, including reports dating back to the 1990s and an online searchable database covering the last ten years.⁴⁵⁰ The MAUDE database allows users to search for medical devices that may have malfunctioned or caused death or serious injury.⁴⁵¹ Users can search by a pull down menu of product problems, by the class of product, or by manufacturer, model, or brand name.⁴⁵² Like FAERS, MAUDE is designed to help the FDA monitor emerging product safety problems, but unlike FAERS, the centralized search function makes the data more accessible to lay users.

Like other agencies, the FDA is trying to publish FAERS, MAUDE, and other databases in more user-friendly formats on its Open FDA site.⁴⁵³ The site, launched in 2014, includes

446. *Id.*

447. *MedWatch: The FDA Safety Information and Adverse Event Reporting Program*, U.S. FOOD & DRUG ADMIN., <http://www.fda.gov/Safety/MedWatch/default.htm> (last updated July 13, 2017) [<https://perma.cc/LA7G-D9YP>]; see also David A. Kessler et al., *Introducing MEDWatch: A New Approach to Reporting Medication and Device Adverse Effects and Product Problems*, 269 JAMA 2765 (June 2, 1993).

448. *MAUDE - Manufacturer and User Facility Device Experience*, U.S. FOOD & DRUG ADMIN., <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm> (last updated June 30, 2017) [<https://perma.cc/GJ8S-LG88>] [hereinafter *MAUDE*].

449. Medical Device Reporting, 49 Fed. Reg. 36,326 (Sept. 14, 1984).

450. *Mandatory Reporting Requirements: Manufacturers, Importers, and Device User Facilities*, U.S. FOOD & DRUG ADMIN., <https://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/ReportingAdverseEvents/ucm2005737.htm> (last updated Nov. 7, 2016) [<https://perma.cc/Y4BF-49CG>].

451. *Id.*

452. *MAUDE*, *supra* note 448.

453. CORTEZ, *supra* note 143, at 44; U.S. Food & Drug Admin., *Open-source APIs and a Developer Community for FDA Data*, OPENFDA, <https://open.fda.gov>

separate pages for food products, drugs, and devices, with all three including enforcement reports, and the drug and device pages, including adverse event databases.⁴⁵⁴ Open FDA publishes both individual reports and larger trend analyses. For example, as of March 2016, the site included almost 5.9 million records in its adverse drug event database dating back to 2004.⁴⁵⁵ Open FDA includes extensive data tools and downloadable raw data files, obviously directed at third-party users.

Also like other agencies, the FDA is beginning to incorporate multiple data sources to pursue regulatory goals, in this case uncovering trends with medical product safety. In 2007, Congress required HHS and the FDA to coordinate with non-FDA sources, including “public, academic, and private entities” to “link and analyze safety data from multiple sources,” with an idea of uncovering emerging product safety risks.⁴⁵⁶ The goal was to include at least 100 million patients in the dataset by 2012.⁴⁵⁷ Called the FDA Sentinel Initiative, the effort has been designed to monitor product safety across different data sources, including data from Medicare, the Veterans Health Administration, and large private health insurers.⁴⁵⁸

But again, like other databases, the FDA’s adverse event databases are not always complete or accurate. A 2011 study found widespread errors and incomplete reports filed in MedWatch, including more than 25 percent of reports using inaccurate product names.⁴⁵⁹ More than most agencies, then,

(last visited July 15, 2017) [<https://perma.cc/QZD4-8SJQ>] [hereinafter *Open-source APIs*].

454. CORTEZ, *supra* note 143, at 44; *Open-source APIs*, *supra* note 453.

455. U.S. Food & Drug Admin., *Adverse Drug Event Reports Since 2004*, OPENFDA, <https://open.fda.gov/drug/event/> (last visited July 15, 2017) [<https://perma.cc/W632-KAXA>].

456. Food and Drug Administration Amendments Act of 2007, Pub. L. No. 110-85, § 905(a), 121 Stat. 823, 944 (codified as amended in scattered sections of 21 U.S.C.).

457. 21 U.S.C. § 355(k)(3)(B).

458. Madison, *supra* note 30, at 1615–16; *see also* Barbara J. Evans, *Authority of the Food and Drug Administration to Require Data Access and Control Use Rights in the Sentinel Data Network*, 65 FOOD & DRUG L.J. 67 (2010) (considering whether participation in the Sentinel Initiative might be mandatory rather than voluntary).

459. Kenneth A. Getz et al., *Evaluating the Completeness and Accuracy of MedWatch Data*, 21 AM. J. THERAPEUTICS 442 (2014) (relying on a 2011 study by the Tufts Center for the Study of Drug Development).

the FDA includes prominent disclaimers of the accuracy and reliability of its data.⁴⁶⁰ For example, the MAUDE database includes the following disclaimer:

Although [Medical Device Reports] are a valuable source of information, this passive surveillance system has limitations, including the potential submission of incomplete, inaccurate, untimely, unverified, or biased data. In addition, the incidence or prevalence of an event cannot be determined from this reporting system alone due to potential under-reporting of events and lack of information about frequency of device use.

* * *

Confirming whether a device actually caused a specific event can be difficult based solely on information provided in a given report. Establishing a cause-and-effect relationship is especially difficult if circumstances surrounding the event have not been verified or if the device in question has not been directly evaluated.⁴⁶¹

Similarly, FAERS emphasizes that the “data does have limitations”:

First, there is no certainty that the reported event (adverse event or medication error) was actually due to the product. FDA does not require that a causal relationship between a product and event be proven, and reports do not always contain enough detail to properly evaluate an event. Further, FDA does not receive reports for every adverse event or medication error that occurs with a product. Many factors can influence whether or not an event will be reported, such as the time a product has been marketed and publicity about an event. Therefore, FAERS data cannot be used to calculate the incidence of an adverse event or medication error in the U.S. population.⁴⁶²

The Open FDA site also confronts users, via pop-up window, with a note of caution that “[t]his API is not for

460. CORTEZ, *supra* note 143, at 44.

461. MAUDE, *supra* note 448.

462. FAERS, *supra* note 441.

clinical or production use. While we make every effort to ensure that data is accurate, you should assume all results are unvalidated.”⁴⁶³ Should policymakers settle for data of such quality? Are incomplete and unrepresentative data better than none? If so, how best can scarce resources be deployed to improve the accuracy, reliability, and usefulness of the FDA’s data?

There remains great hope that modern tools like agency databases can improve drug safety, even if the data are imperfect, by combining FDA data with other data sources, like Medicare claims, for example.⁴⁶⁴ Thus, an alliance between the FDA and CMS to combine their data might help cure some defects in the FDA’s adverse event reporting systems.⁴⁶⁵ Still, policymakers might counter widespread underreporting by experimenting with automated monitoring systems relying on digital technologies. For example, digital pill trackers and other mobile monitoring devices might send automatic problem reports to the FDA,⁴⁶⁶ subject to pre- and post-publication safeguards described above.

V. DESIGNING FOR OPTIMAL DISCLOSURE

How can policymakers design databases that realize their many aspirations while minimizing their shortcomings and burdens? At core, how can policymakers ensure the quality and reliability of agency data, so that users *trust* the data being published?⁴⁶⁷ My first prescription is modesty. As Richard Craswell cautions, “people who expect disclosure laws to solve

463. See, e.g., U.S. Food & Drug Admin., *Drugs – Adverse Events*, OPENFDA, <https://open.fda.gov/drug/event/> (last visited July 16, 2017) [<https://perma.cc/4JR2-TB5R>].

464. See, e.g., Scott Gottlieb, *Opening Pandora’s Pillbox: Using Modern Information Tools to Improve Drug Safety*, 24 HEALTH AFF. 938 (2005).

465. *Id.*

466. For a very broad survey of some of the recent digital technologies, see Nathan Cortez, *The Mobile Health Revolution?*, 47 U.C. DAVIS L. REV. 1173 (2014).

467. These values derive, not coincidentally, from the Information Quality Act. Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515, Pub. L. No. 106-554, § 515, 114 Stat. 2763, 2763A-153-54 (2001); 44 U.S.C. § 3516. Note also that the GAO has discussed what it means for data to be “reliable.” See APPLIED RESEARCH & METHODS, U.S. GOV’T ACCOUNTABILITY OFFICE, ASSESSING THE RELIABILITY OF COMPUTER PROCESSED DATA 5 (2009), <http://www.gao.gov/assets/80/77213.pdf> [<https://perma.cc/DR5L-XEER>].

almost every problem—quickly, easily, and with very little cost—are doomed to have their expectations crushed.”⁴⁶⁸ The truth is that regulation by database requires just as many difficult design and implementation choices as any other form of regulation.⁴⁶⁹ The early successes of the CFPB’s Consumer Complaint Database and the CPSC’s SaferProducts.gov site, for example, are owed to several canny decisions by Congress and the agencies to ensure the quality of the data posted.⁴⁷⁰ It is a mistake, then, to assume that creating a public database is necessarily less difficult and more cost-effective than traditional regulation.

Another important consideration is that agency databases vary widely in their purposes, scope, design, sources, and presentation. Although it is neither possible nor worthwhile to prescribe universal rules of thumb here, I try to highlight emerging best practices from the databases I have evaluated to date, in the hopes that these discussions will be useful to policymakers. Again, these recommendations build on, and in many ways exceed, those recently promulgated by ACUS.⁴⁷¹ In short, the most successful disclosure regimes will carefully consider both data inputs (how data will be collected and from whom) and outputs (how the data will be published and presented).⁴⁷²

I thus offer a series of recommendations that address both dimensions, proposing a decidedly modern role for the government as a “data steward.” For example, agencies must consider how to gather and process the data, including carefully choosing sources and crafting adequate pre- and post-publication procedures to ensure data quality. Agencies also should not underestimate the costs required to generate and maintain their databases, as well as the potential overlap with agency enforcement and adjudication procedures. Finally, agencies should think carefully about potential audiences and users in deciding how to characterize and present the data.

468. Craswell, *supra* note 18, at 379.

469. *Id.*

470. See discussion *supra* Sections IV.C & IV.D.

471. ACUS Adoption of Recommendations, 81 Fed. Reg. 40,259 (June 21, 2016).

472. For parallel thinking on how to improve quality reporting in health care, including data selection, design, reporting, and presentation, see Kristin Madison, *Legal and Policy Issues in Measuring and Improving Quality*, in THE OXFORD HANDBOOK OF U.S. HEALTH LAW 680 (I. Glenn Cohen et al. eds., 2016).

A. Gathering and Processing the Data

Policymakers should think carefully, and ideally in advance, about data inputs. Whose data will populate the data sets? How will it be gathered? To what extent will the agency try to verify, validate, or otherwise authenticate the data? And how will the agency handle contested data?

As a threshold matter, it is particularly important that agencies identify reliable data sources. The irony here is that the internet not only facilitates many of the disclosures discussed in this Article, but also (by virtue of soliciting data from a variety of sources) raises problems with reliability.⁴⁷³ Thus, government agencies can play an important role by ensuring that published data is credible.⁴⁷⁴ Indeed, because the very fact of publication by a government agency often signals credibility, it is incumbent on agencies to ensure the accuracy of data that purports to be objective.⁴⁷⁵ Notice-and-comment procedures might be useful in soliciting feedback on what sources are reliable and why.

As Kristin Madison argues, the federal government is more than just a repository for data—it is also a “data steward” responsible for actively managing the data it holds, helping to ensure its integrity.⁴⁷⁶ The CFPB, for example, does not verify that consumer complaints are “accurate” (in other words, that the conduct alleged in the complaint actually occurred), but does help “authenticate” that each complaint is made by an actual customer of the company, giving the company ample opportunity to identify false or fraudulent complaints.⁴⁷⁷ Likewise, when Congress authorized SaferProducts.gov, it required the CPSC to consider objections that the information is “materially inaccurate.”⁴⁷⁸ Again, the CFPB and CPSC demonstrate two frameworks for active data stewardship.

In short, agencies that purport to publish accurate and objective data should adopt procedures to these ends. As

473. See, e.g., FUNG ET AL., *supra* note 15, at 165 (“Those who contribute information can do so without identifying themselves or their sponsoring organizations, or taking responsibility for what they are saying.”).

474. Heerwig & Shaw, *supra* note 243, at 1476.

475. CORTEZ, *supra* note 143, at 44.

476. Madison, *supra* note 30, at 1607–08.

477. Disclosure of Consumer Complaint Data, 78 Fed. Reg. 21,218, 21,225–26 (Apr. 10, 2013).

478. 15 U.S.C. § 2055a(c)(4) (2012).

emphasized in Part II, because “transparency” is frequently invoked to support disclosure, it is important to remember that “transparency” allows objects to be “seen without distortion.”⁴⁷⁹ To the extent feasible, then, government databases should try to achieve genuine transparency rather than “translucency.”⁴⁸⁰ Indeed, some scholars note that, inevitably, disclosure regimes created through the political process are forged by conflict and compromise, and thus generate only partial or imperfect transparency.⁴⁸¹ But scholars also find that the more successful disclosure regimes tend to increase the accuracy and quality of the information they publish over time.⁴⁸² Databases should be dynamic rather than static—constant works in progress. Although initial design choices are important, agencies should not hesitate to tinker with data collection procedures if flaws in the data become apparent, as in the case of the EPA’s Toxic Release Inventory.

1. Pre-Publication Procedures

Data sets that purport to publish accurate, objective information should be buttressed, ideally, with both pre- and post-publication procedures that allow parties to request that any information not meeting these standards be corrected or retracted. Again, some of the newer agency databases provide parties with pre-publication procedures to comment on, challenge, or request corrections and retractions of information before publication.⁴⁸³ For example, by statute the CPSC must give companies whose products are reported to SaferProducts.gov the opportunity to comment on any Reports of Harm.⁴⁸⁴ The CPSC must consider objections that the information is “materially inaccurate,”⁴⁸⁵ and the Commission publishes these procedures in the C.F.R.⁴⁸⁶ There are clear

479. Schauer, *supra* note 9, at 1343 (referring to the *Oxford English Dictionary*’s definition of “transparency”).

480. *Id.* at 1345. The Schauer distinction between transparency and translucency is subject to the objection, of course, that it assumes that the state could ever be completely “transparent,” providing undistorted access to itself.

481. FUNG ET AL., *supra* note 15, at xii.

482. *Id.* at 109.

483. CORTEZ, *supra* note 143, at 94–95.

484. 15 U.S.C. § 2055a(c)(2) (2012).

485. *Id.* § 2055a(c)(4).

486. 16 C.F.R. § 1102.26.

timelines for parties to object to alleged inaccuracies and for the CPSC to resolve disputes before publication.⁴⁸⁷ Likewise, the CFPB authenticates that complaints are coming from actual customers of the company.⁴⁸⁸ Bureau procedures allow companies to use an online company portal to verify a commercial relationship with the customer and post the company's response.⁴⁸⁹ The Bureau also makes clear that each complaint, before being published in the database, must meet several publication criteria.⁴⁹⁰ Of all the databases discussed in Part IV, the CPSC and CFPB procedures serve as the best models for pre-publication quality control.

2. Post-Publication Procedures for Contested Data

Nevertheless, errors in published data sets are probably inevitable, no matter how robust the pre-publication procedures may be. As a result, policymakers should also consider post-publication procedures as a backstop to help ensure the quality and reliability of data. Scholars have long recognized that such procedures can be an important safety valve for parties named in agency publications, as legal recourse is generally not available.⁴⁹¹

Here, the controversial Information or Data Quality Act might be of help.⁴⁹² The Act required the OMB to publish government-wide guidelines for “ensuring and maximizing the quality, objectivity, utility, and integrity of information . . . disseminated by the government.”⁴⁹³ The Act applies broadly to “[f]ederal agency dissemination of public information, regardless of the form or format.”⁴⁹⁴ It also directed the OMB to

487. 15 U.S.C. § 2055.

488. Disclosure of Consumer Complaint Data, 77 Fed. Reg. 37,616, 37,616–17 (June 22, 2012).

489. COMPANY PORTAL MANUAL, *supra* note 316.

490. CORTEZ, *supra* note 143, at 63–64.

491. Gellhorn, *supra* note 169; CORTEZ, *supra* note 143, at app. C.

492. The Act has been criticized as an attempt to thwart regulators in the guise of data quality. *See, e.g.*, Levy & Johns, *supra* note 27, at 2.

493. Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515, Pub. L. No. 106-554, § 515, 114 Stat. 2763, 2763A-153-54 (2001). The IQA is built on earlier provisions in the Paperwork Reduction Act. Paperwork Reduction Act of 1995, Pub. L. No. 104-13, 109 Stat. 163, 168.

494. 67 Fed. Reg. 8452, 8460 (Feb. 22, 2002). OMB guidelines define “information” as “any communication or representation of knowledge such as facts

establish procedures that allow “affected persons to seek and obtain correction of information maintained and disseminated by the agency.”⁴⁹⁵ Per the OMB’s guidelines, dozens of federal agencies have published their own such guidelines and post-publication procedures for correcting or retracting information.⁴⁹⁶ Although these procedures would seem to have clear application to agency databases, the OMB guidelines include two important exemptions. First, they exclude from the IQA’s coverage “opinions, where the agency’s presentation makes it clear that what is being offered is someone’s opinion rather than fact or the agency’s views.”⁴⁹⁷ Second, they exclude “adjudicative processes.”⁴⁹⁸ Thus, both exemptions could be read as excluding, for example, the CFPB’s Consumer Complaint Database, which might be fairly characterized as including “opinions” or even “adjudicative processes.”⁴⁹⁹ Nevertheless, the IQA and resulting agency guidelines articulate both substantive and procedural values that agencies should observe. To ensure the quality and reliability of government-published information, there should be a safety valve that allows the subjects identified to request correction or retraction by the agency.

A 2015 study by the GAO found eighty-seven publicly-reported requests for corrections or retractions under the IQA sent to thirty agencies between 2010 and 2014.⁵⁰⁰ Although the agencies denied fifty-nine out of the eighty-seven requests, the agencies made full corrections in eleven cases and partial

or data, in any medium or form,” including “information that an agency disseminates from a web page.” *Id.*

495. Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515.

496. See *Agency Information Quality Guidelines*, OFFICE OF MGMT. & BUDGET, https://obamawhitehouse.archives.gov/omb/infoleg_agency_info_quality_links/ [<https://perma.cc/8YZZ-PKRF>].

497. 67 Fed. Reg. at 8453–54.

498. *Id.* at 8454.

499. CORTEZ, *supra* note 143. Of course, industry members have argued that the CFPB’s database is subject to the IQA. See, e.g., Letter from Wayne A. Abernathy of the American Bankers Association to Hon. Mark Bialek, Inspector General, Board of Governors of the Federal Reserve System (Jan. 12, 2015), <http://www.cfpbmonitor.com/files/2015/01/LTC-ConsCompDatabase2015Jan.pdf> [<https://perma.cc/G8JA-XWBL>].

500. U.S. GOV’T ACCOUNTABILITY OFFICE, GAO-16-110, INFORMATION QUALITY ACT: ACTIONS NEEDED TO IMPROVE TRANSPARENCY AND REPORTING OF CORRECTION REQUESTS 8 (2015) <http://www.gao.gov/assets/680/674386.pdf> [<https://perma.cc/4KXN-FHHF>] [hereinafter GAO-16-110].

corrections in fifteen cases (two cases were unresolved as of the date of the report).⁵⁰¹ For example, HHS updated information on several CDC websites regarding bicycle helmet safety data in response to a request made under CDC, HHS, and OMB data quality guidelines.⁵⁰² Moreover, in fifteen of the fifty-nine cases denying the IQA request, the agency used alternative procedural mechanisms, usually systems that predated the IQA.⁵⁰³ In some cases, the agency engaged in long substantive exchanges on the accuracy, presentation, and usability of data.⁵⁰⁴

Post-publication procedures might also reside outside, rather than inside, the agency. My review for ACUS, for example, considered whether independent bodies like the OMB, ombudsmen, or inspectors general might play a role in supervising disputes over agency data.⁵⁰⁵

First, the OMB already exerts both centripetal and centrifugal pressures on agency data collection and publication practices. In addition to the OMB's IQA guidelines, the Paperwork Reduction Act requires the OMB to preapprove significant information gathering efforts by agencies.⁵⁰⁶ The agency must explain to the OMB why it needs the information, why it has "practical utility," and why it is relevant to the agency's regulatory functions.⁵⁰⁷ Thus, the OMB can play a useful standardizing role. But it is not well suited to resolving disputes between agencies and regulated parties, and agencies may bristle at having to endure further layers of OMB review.

Second, many agencies maintain an Office of the Ombudsman or its equivalent, which can field complaints about data published by agencies. For example, the CFPB's Ombudsman has heard complaints about the Consumer

501. *Id.* at 17–18 tbl.4.

502. Letter from James G. Titus, Washington Area Bicyclist Ass'n, to Ctr. for Disease Control (Feb. 28, 2013), <https://aspe.hhs.gov/system/files/pdf/105661/44aCDC.pdf> [<https://perma.cc/7GQG-ANX6>]; Response Letter from G. David Williamson, Ctr. for Disease Control, to James G. Titus, Washington Area Bicyclist Ass'n (July 11, 2013), <https://aspe.hhs.gov/cdc---bicycle-helmet-safety-b2> [<https://perma.cc/5UGN-8U44>].

503. GAO-16-110, *supra* note 500, at 18–21.

504. *See, e.g.*, Letter from Mary Pratt, CMS, to Jackson Williams, Dir. of Gov't Affairs, Dialysis Patient Citizens (Sept. 9, 2015), <https://aspe.hhs.gov/system/files/pdf/110656/51b5.pdf> [<https://perma.cc/LJA9-2NGX>].

505. CORTEZ, *supra* note 143, at 100–02.

506. 44 U.S.C. § 3507(d) (2012).

507. *Id.* § 3508.

Complaint Database,⁵⁰⁸ and the FDA's many ombudsmen have fielded complaints under the IQA.⁵⁰⁹ Ombuds can serve important customer service functions with regulated parties. As such, the use of ombuds in federal agencies has increased in recent years, as have calls for standards regarding their independence, duties, and information-providing roles.⁵¹⁰

A third option is review by agencies' inspectors general. Inspectors general (IGs) are independent officers, directed by law to detect and prevent fraud, waste, and abuse in federal agencies.⁵¹¹ They also maintain, by design, crucial independence from agency heads, and thus can serve as an independent arbiter.⁵¹² They can also function as an avenue for fielding industry complaints.⁵¹³ The Federal Reserve's Office of Inspector General is, in fact, auditing the CFPB's Consumer Complaint Database "to assess the effectiveness of the CFPB's controls over the accuracy and completeness of the public complaint database."⁵¹⁴

Finally, chief information officers (CIOs) within agencies might play an important role in not only answering important questions regarding database design, but also in participating in pre- and post-publication procedures described above. They are most likely to be informed of other agencies' experiences and able to critically evaluate whether those models might translate well to their own data regimes.

508. CONSUMER FIN. PROT. BUREAU, OMBUDSMAN'S OFFICE, ANNUAL REPORT TO THE DIRECTOR 5, 6 (2014).

509. OFFICE OF THE OMBUDSMAN, U.S. FOOD & DRUG ADMIN., DISPUTE RESOLUTION AND PROBLEM SOLVING 2, <http://oig.federalreserve.gov/reports/work-plan-full.htm#CFPBongoing> (last visited Oct. 20, 2017) [<https://perma.cc/BQ2X-PF2T>].

510. See, e.g., ADMIN. CONFERENCE OF THE U.S., *The Use of Ombuds in Federal Agencies*, (Dec. 14, 2016), <https://www.acus.gov/recommendation/use-ombuds-federal-agencies> [<https://perma.cc/9QT5-9ZAK>].

511. Inspector General Act of 1978, Pub. L. No. 95-452, 92 Stat. 1101.

512. See, e.g., Neal Kumar Katyal, *Internal Separation of Powers: Checking Today's Most Dangerous Branch from Within*, 115 YALE L.J. 2314, 2347 (2006).

513. CORTEZ, *supra* note 143, at 101.

514. *Work Plan: Audit of the CFPB's Public Consumer Complaint Database*, OFFICE OF INSPECTOR GEN., FED. RESERVE, <https://oig.federalreserve.gov/reports/cfpb-management-controls-consumer-complaint-database-sep2015.pdf> (last visited June 5, 2015) [<https://perma.cc/95SH-V8TQ>].

3. Considering Costs

Procedural safeguards can be essential for ensuring data quality, though they are not without cost. Too often government agencies try to achieve disclosure on the cheap.⁵¹⁵ Data collection and processing requires not just automation, but also human labor. Unfortunately, such labor is dismissed as the task of “data janitors” who receive inadequate compensation.⁵¹⁶ The lack of sufficient infrastructure to ensure data quality can generate “big bad data”⁵¹⁷—data that are voluminous but of low quality. Thus, meaningful data often require meaningful investment to create a sufficient information infrastructure.

Moreover, posting the data online can be costly. Running an agency website is not a simple proposition. Web masters for federal websites must comply with at least two dozen different regulatory systems “[r]anging from privacy and usability to FOIA compliance to the demands of the Paperwork Reduction Act.”⁵¹⁸ Although each separate requirement may stand on its own logic, together they can limit how agencies present data, and generally favor standardization above experimentation.⁵¹⁹ As Robinson and colleagues observe, “[a]s long as the government has a special role in the presentation and formatting of raw government data, certain desirable limits on what the government can do become undesirable limits on how the data can be presented or handled.”⁵²⁰ In this vein, nongovernmental intermediaries have proven useful in rendering government data more accessible and usable,⁵²¹ as I discuss below in Section V.C.

Finally, all this assumes that policymakers have already made the threshold decision to publish the data and accept

515. I am indebted to Frank Pasquale for raising many of the issues in this paragraph.

516. See, e.g., Lilly Irani, *Justice for “Data Janitors,”* (Jan. 15, 2015), <http://www.publicbooks.org/nonfiction/justice-for-data-janitors> [<https://perma.cc/YK9H-6AYU>].

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

518. David Robinson et al., *Government Data and the Invisible Hand*, 11 YALE J.L. & TECH. 160, 162 (2009).

519. *Id.* at 163–65.

520. *Id.* at 165.

521. *Id.*

responsibility for data stewardship. But because data stewardship can be resource intensive, policymakers should think more critically in advance about *which* data regimes warrant the government's scarce stewardship resources. Which databases might be especially useful to consumers, or particularly effective at inducing optimal behavior from regulated entities? And of existing databases, which low-quality data sets might be worth salvaging? The FEC's and FDA's databases might underwhelm in several ways, but few would argue that more accurate, comprehensive campaign finance data or drug safety data would not be worthwhile. Thus, if stewardship is worth pursuing, limited resources dictate that it must be *targeted* stewardship, focusing on the data sets that are most likely to achieve the twin goals of being useful to target audiences and changing behavior.⁵²²

4. Administrative Law Dimensions

Most interesting, from an administrative law perspective, is that database publication procedures might be viewed as a unique species of agency adjudication and regulatory enforcement. Seen this way, database publication procedures fall along a continuum ranging from very minimal verification to more searching pre- and post-publication adjudication of contested data. No database regimes currently approach the full panoply of procedural safeguards that attach to more formal administrative adjudications, pursuant to statutory and due process requirements.⁵²³ But like traditional agency adjudications, the amount of "procedure" appropriate for database disputes will depend on what information each database includes, the regulatory goals of publishing the information, the statutory scheme in which it sits, and the cost and value of "getting it right" versus "making it public."

Another possibility is that legal sanctions begin to attach more formally and more forcefully to database reporting such that companies reporting inaccurate or incomplete data will face fines, penalties, and other measurable burdens in addition to any reputational damage. Could the Medicare program, for example, condition reimbursement on accurate data

522. I credit Nick Bagley for seeding the ideas in this paragraph.

523. Bill Sage inspired many of the thoughts in this paragraph.

reporting?⁵²⁴ Could plaintiffs or prosecutors use the federal false statements statute⁵²⁵ or the False Claims Act⁵²⁶ to punish material inaccuracies or misleading data reporting by regulated firms? Both laws are broad and powerful, and are deployed in increasingly creative ways. Their use in database reporting cases could raise novel statutory and due process questions about the procedural safeguards agencies have selected for specific databases. Moreover, such actions could be undermined by how agencies themselves “characterize” the data, including any disclaimers about the accuracy, reliability, or objectivity of the data.

B. Characterizing the Data

Because not all data will be objective in nature, verifiable, or even worth the cost of verifying, it is equally important that agencies accurately characterize the data they present, listing the sources and any important context or limitations for the data.

1. Identifying Sources

Databases should be labeled and characterized accurately, much as we expect of product labeling by industry.⁵²⁷ This is particularly so because government agencies are one of the most trusted sources of information, and the information they publish carries the imprimatur of the federal government.⁵²⁸ Thus, agency databases should clearly identify the sources of their data. Agency databases should also indicate whether the data are contested, and detail steps the agency takes (or, more importantly, does not take) to resolve such contests.⁵²⁹

524. For a discussion of this possibility, see Frank Pasquale, *Grand Bargains for Big Data: The Emerging Law of Health Information*, 72 MD. L. REV. 682 (2013).

525. 18 U.S.C. § 1001 (2012). Kristin Madison discusses the possibility of prosecutors using the False Claims Act to punish misreporting of quality data in pay-for-performance or value-based purchasing programs under Medicare. See Madison, *supra* note 472, at 692.

526. False Claims Act, 31 U.S.C. §§ 3729–33 (2012).

527. See, e.g., 21 U.S.C. § 343(a) (2012) (prohibiting food labeling from being false or misleading).

528. CORTEZ, *supra* note 143, at 94.

529. *Id.* at 96.

Data sets that do not purport to be accurate or objective might require special precautions. Federal databases can be populated with data from a variety of sources—including consumers, regulated parties, or the agency itself—and each might require different quality controls and presentations.⁵³⁰ For example, the FDA’s adverse event databases are populated by reports from manufacturers and users that a product may have been “associated” with an adverse event, without any firm claims as to causation.⁵³¹ Similarly, being listed in the CFPB’s Consumer Complaint Database does not mean that a company has committed any *legal* violation; many complaints are simply “vague expressions of being wronged.”⁵³² Just like the FDA does not verify whether a product caused a specific adverse event, the CFPB does not verify that a company even engaged in the conduct alleged in the consumer complaint. Doctors subject to “report cards” also lament that death and complication rates are presented without being normalized for treating riskier patient populations.⁵³³ Should such “data” even be published? For better or worse, routinely they are.

2. Explaining Context, Limitations

The solution, perhaps, is for agencies to represent the data accurately, which often means explaining the context and any limitations of the data. Several agencies already endeavor to do so. For example, in the FDA’s medical device database, the agency notes that its “surveillance system has limitations, including the potential submission of incomplete, inaccurate, untimely, unverified, or biased data.”⁵³⁴ The FDA’s adverse drug event database also notes that “there is no certainty that the reported event . . . was actually due to the product.”⁵³⁵ Likewise, the CFPB disclaims that “[w]e don’t verify all the facts alleged in these complaints but we take steps to confirm a

530. *Id.*

531. See discussion *supra* Section IV.F.

532. CORTEZ, *supra* note 143, at 69–70; Porter, *supra* note 163, at 78.

533. Sandeep Jauhar, Opinion, *Giving Doctors Grades*, N.Y. TIMES (July 22, 2015), <http://www.nytimes.com/2015/07/22/opinion/giving-doctors-grades.html> [perma.cc/WH3T-Q6NJ]. Although death and complication rate data are usually normalized through various methodologies, physicians frequently object that such methodologies are inadequate.

534. MAUDE, *supra* note 448.

535. FAERS, *supra* note 441.

commercial relationship between the consumer and the company.”⁵³⁶ Congress requires the CPSC’s SaferProducts.gov database to “provide clear and conspicuous notice to users of the database that the Commission does not guarantee the accuracy, completeness, or adequacy of the contents of the database.”⁵³⁷ Although one court called the CPSC’s language “boilerplate” that “would not interest an ordinary consumer,”⁵³⁸ providing appropriate context and disclosing the limitations of data is relatively easy and helps answer several criticisms of disclosure noted in Part III.

A “reliable” database may depend not only on publishing accurate data (if that is what it purports to do), but also on publishing relatively complete and representative data. Industry commenters, for example, objected that the CFPB’s database of self-selected consumer complaints would necessarily be incomplete, nonrandom, and thus nonrepresentative of company performance and consumer experiences.⁵³⁹ The CFPB responded that the data are not portrayed as such, and promised to “inform consumers and any other public database users that the data reflect only the . . . complaints that consumers submit to the Bureau.”⁵⁴⁰

Data selection or filtering criteria might thus generate published data that are technically accurate but misleading as a whole. Transparency initiatives often fail when “transparency is either not sufficiently mandatory or not applicable to categories of information that meaningfully contribute to public accountability.”⁵⁴¹ Agencies with discretion to disclose information may tend to disclose “information that makes the administration look public spirited, effective, and efficient, but withhold information to the contrary.”⁵⁴² The data chosen for publication may not paint a complete or representative picture. Thus, as Shkabatur argues, the answer for incomplete transparency may be even more

536. *Consumer Complaint Database*, *supra* note 312.

537. 15 U.S.C. § 2055a(b)(5) (2012).

538. *Co. Doe v. Tenenbaum*, 900 F. Supp. 2d 572, 598 (D. Md. 2012), *rev’d sub nom.* *Co. Doe v. Public Citizen*, 749 F.3d 246 (4th Cir. 2014).

539. *See, e.g.*, 77 Fed. Reg. 37,616 (June 22, 2012).

540. *Id.*

541. Shkabatur, *supra* note 55, at 105–06.

542. Adam Samaha, *Government Secrets, Constitutional Law, and Platforms for Judicial Intervention*, 53 UCLA L. REV. 909, 919 (2006); Shkabatur, *supra* note 55, at 106.

transparency.⁵⁴³ She finds support among scholars who argue that Congress should place affirmative disclosure duties on agencies, shifting away from the “passive” disclosure required by laws like FOIA that have been rendered as anachronisms in the Internet era.⁵⁴⁴ If it is neither possible nor cost-effective to publish comprehensive or representative data, the agency should provide adequate context for what is being published and explain why the dataset is incomplete.

C. Presenting the Data

Policymakers should also consider the “outputs” of agency databases—how the data will be published, presented, and used. Thoughtful designs will evaluate the optimal format, size, and scope of the database, as well as the target audiences and their potential uses.

1. Raw or Polished?

First, data can be published in raw or relatively polished formats, with gradations in between. Should agencies rely on massive raw data dumps targeted at more sophisticated users? Or should they package, stylize, and distill the data for lay users? The trend, as noted above, is to publish data sets in both more polished, packaged formats and in raw, open data formats. The former requires agencies to think carefully about how to convey the information, and in what packaging—which inevitably raises costs and includes normative judgment calls that might draw into question how objective or neutral the presentation is. But the latter (publishing raw data) is a relatively recent phenomenon.

Historically, agencies have been reluctant to publish information in open, raw, machine-readable data formats (particularly information requested via FOIA). For example, 2011 congressional testimony revealed that “[m]ost requests for correspondence and other documents are fulfilled by printing them, redacting, then re-scanning into unsearchable

543. See generally Shkabatur, *supra* note 55.

544. See, e.g., Vladeck, *supra* note 64, at 1828–29.

images.”⁵⁴⁵ Yet, as far back as 2004, the OMB encouraged agencies to “provide all data in an open, industry standard format permitting users to aggregate, disaggregate, or otherwise manipulate and analyze the data to meet their needs.”⁵⁴⁶ And scholars continue to argue for agencies to publish data online in open, structured, and machine-readable formats such as XML, consistent with the Open Government Working Group’s recommendations.⁵⁴⁷ Thus, there is a clear trend toward publishing in raw, open formats.

At the same time, some scholars argue that publishing data in raw, “naked” formats can itself serve as a barrier to access for nonprogrammers and others who are not able to understand or use such data.⁵⁴⁸ Raw government datasets might require, ironically, technically sophisticated intermediaries to decipher.⁵⁴⁹ Thus, open government efforts that encourage agencies to present data in a raw, naked, and “neutral” way may erect separate barriers to accessing and understanding the information.

A related idea is that government transparency can exist on different planes, from “relative” to “absolute.” Relative transparency occurs when someone—usually the government or a data intermediary—“relates” the data from one reporting entity to another for easy comparison. For example, star ratings, grades, and other distilling criteria essentially grade reporting entities on a curve. To wit, most users would understand that a hospital receiving one out of five stars on overall quality underperforms most other hospitals. But it is

545. *The Freedom of Information Act: Ensuring Transparency and Accountability in the Digital Age: Hearing Before the S. Comm. on the Judiciary*, 112th Cong. (2011) (statement of Sarah Cohen, Knight Professor of the Practice of Journalism, Duke University), <https://www.judiciary.senate.gov/imo/media/doc/11-3-15%20Cohen%20Testimony.pdf> [<https://perma.cc/W7PM-9SNM>].

546. OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT, OMB MEMORANDUM M-05-04, POLICIES FOR FEDERAL AGENCY PUBLIC WEBSITES 4 (2004), <https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2005/m05-04> [<https://perma.cc/B6PN-37HC>].

547. See, e.g., Robinson et al., *supra* note 518, at 167 (arguing that original data should be posted in documents in XML formats with unique and permanent addresses); *The Annotated 8 Principles of Open Government Data*, OPEN GOV’T WORKING GRP., <http://opengovdata.org/index.php/OpenDataPrinciples> (last visited Oct. 8, 2017) [<https://perma.cc/5YEA-HMJ6>] (recommending that data be complete, primary, timely accessible, machine-readable, non-discriminatory, non-proprietary, and license-free).

548. See, e.g., Shkabatur, *supra* note 55, at 112.

549. *Id.*

much more difficult for users to understand what a two percent complication rate associated with cardiac surgeries performed at a specific facility should signal, as an absolute number. Of course, even “relative” transparency may not be particularly useful. If HospitalCompare.gov lists a hospital’s mortality rates as “[n]o different than the national average,” that might signal to users that they should not worry about that factor when selecting a hospital. But it could also mean that the national average is equally disappointing for everyone. Moreover, “relative” transparency is only realized after gathering “absolute” data points. But who should take on the task of turning absolute, raw data into relative, packaged comparisons?⁵⁵⁰

Some argue that the government should focus its energies less on presenting packaged information and more on publishing “reusable data.”⁵⁵¹ The idea, inspired by the engineering principle that separates data from interaction, is that agencies should worry less about designing user-friendly websites, and more about releasing raw data for nongovernmental users.⁵⁵² Robinson and colleagues argue that the latter will be better able to experiment with how to present the data effectively, whether it be with advanced search functionalities, automated content analysis, indexing among multiple sources, and various data visualization tools.⁵⁵³ They call this new role for agencies an “invisible hand,” enabling a “marketplace of engineering ideas.”⁵⁵⁴ Some users will also value being able to access “genuine” data that is not mediated, framed, or translated by an intermediary (including, or even particularly, by the government).⁵⁵⁵

550. I credit Kristin Madison with raising the notion of “relative” versus “absolute” transparency, which parallels some of the considerations when deciding whether to prioritize publishing raw versus polished data sets. Larry Lessig also touches on this dynamic when comparing campaign finance data (absolute) with EPA fuel economy stickers on new cars (relative). See LESSIG, *supra* note 260, at 257–58.

551. Robinson et al., *supra* note 518, at 160.

552. *Id.* at 161.

553. *Id.* at 161, 169.

554. *Id.* at 161.

555. *Id.* at 174. Fenster discusses how the process of the government communicating information necessarily involves imperfect judgment calls about what information to disclose and how. Fenster, *supra* note 26, at 926–27.

2. Big Data or Small Data?

Disclosure enthusiasts often assume that more is better. But recently, scholars have begun to acknowledge that it is more important that information be accessible and *usable*, rather than simply available.⁵⁵⁶ Perhaps the relevant question, then, is not what policymakers think users *need* to know, but what users *want* to know.⁵⁵⁷ Disclosure policies that consider what information users want, and what they can comprehend, tend to be more successful over time.⁵⁵⁸

Moreover, perhaps consumers do not necessarily need more data, but more advice.⁵⁵⁹ The opposite of making data “bigger,” of course, is making data “smaller”—usually by simplifying, tailoring, and targeting the information to make it easier to process.⁵⁶⁰ Thus, rating systems and other information made available at the point of purchase could be particularly useful for consumers.⁵⁶¹ Mere publication on a government website might not be particularly useful, unless intermediaries make the data available *where* and *when* it can be used. Such information is more likely to become embedded in the decisions targeted by the disclosure.⁵⁶² Thus, databases that allow users to simplify complex information, or that allow experts to easily convert it to actionable advice (such as a ratings system or a reliable heuristic), will be more successful in achieving regulatory goals.⁵⁶³ Restaurant hygiene grades, for example, are more embedded in the decision of where to eat than complex and voluminous patient safety disclosures are in the decision of where to seek medical care.⁵⁶⁴

Despite the current fascination with “big data,” many also appeal for simplification. Agencies are thinking more carefully today about ensuring the “utility” of data, perhaps owing in

556. See, e.g., Schauer, *supra* note 9, at 1344; Fenster, *supra* note 26, at 942.

557. Ben-Shahar & Schneider, *supra* note 17, at 746.

558. FUNG ET AL., *supra* note 15, at 11.

559. Ben-Shahar & Schneider, *supra* note 17, at 746.

560. Madison, *supra* note 30, at 1621–22 (noting astutely that recent laws like the Affordable Care Act include language like “patient-centered” and “patient engagement,” which can be code for making big data smaller).

561. FUNG ET AL., *supra* note 15, at 57.

562. *Id.* at 65–74.

563. *Id.* at 57.

564. *Id.* at 65.

part to the IQA.⁵⁶⁵ Moreover, notions of “utility” continue to evolve. For example, the Nutrition Facts label on food products has had some modest success—consumers report that they are increasingly aware of nutrition labeling and make decisions based on it.⁵⁶⁶ Still, as Ben-Shahar and Schneider note, even with nutrition labeling, which they call “the simplest and most understandable case of daily disclosures,” studies still find high levels of consumer confusion that largely correlate with low consumer literacy and numeracy.⁵⁶⁷ How much should these findings deter agency disclosure efforts?

In general, ratings systems that communicate data that has been simplified and “translated” for lay users seem to enjoy moderate success.⁵⁶⁸ For example, there is evidence that simple letter grades for restaurant sanitation (from “A” to “C”) have led to cleaner restaurants in Los Angeles County.⁵⁶⁹ Restaurants with high letter grades posted in their store windows saw an increase in revenues, and conversely, restaurants with the lowest “C” grades saw a decrease.⁵⁷⁰ More tellingly, prominent disclosure of these grades encouraged restaurants to improve their sanitation practices, which correlated with a significant local drop in hospitalizations related to food-borne illnesses.⁵⁷¹ Thus, simple, comprehensible, and easily accessible ratings not only allowed

565. The IQA required the OMB to issue government-wide guidelines for “ensuring and maximizing the quality, objectivity, utility, and integrity of information . . . disseminated by federal agencies.” Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515, Pub. L. No. 106-554, 114 Stat. 2763, 2763A-153-54 (2001); 44 U.S.C. § 3516.

566. Ben-Shahar & Schneider, *supra* note 17, at 675.

567. *Id.* at 675–76 (citing Gary Jones & Miles Richardson, *An Objective Examination of Consumer Perception of Nutrition Information Based on Healthiness Ratings and Eye Movements*, 10 PUB. HEALTH NUTRITION 238 (2007); Gill Cowburn & Lynn Stockley, *Consumer Understanding and Use of Nutrition Labeling: A Systematic Review*, 8 PUB. HEALTH NUTRITION 21, 23 (2005); Russell L. Rothman et al., *Patient Understanding of Food Labels: The Role of Literacy and Numeracy*, 31 AM. J. PREVENTATIVE MED. 391, 391 (2006)).

568. FUNG ET AL., *supra* note 15, at 743.

569. Ben-Shahar & Schneider, *supra* note 17, at 743; Ginger Zhe Jin & Phillip Leslie, *The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards*, 118 Q.J. ECON. 409 (2003) (studying sanitation grades in Los Angeles County).

570. FUNG ET AL., *supra* note 15, at 50 (citing studies).

571. *Id.* Note, however, that studies of other local restaurant grading efforts have found them to be less successful. *See, e.g.*, Daniel E. Ho, *Fudging the Nudge: Information Disclosure and Restaurant Grading*, 122 YALE L.J. 574 (2012) (examining similar efforts in multiple cities).

consumers to vote with their wallets, but also encouraged restaurants to compete based on cleanliness—undoubtedly the underlying motivation of the letter grade system.⁵⁷²

For disclosure policies to succeed on multiple levels, then, they must affect not only the decision-making of consumers and regulatory beneficiaries, but also the decision-making of the discloser—the regulated party.⁵⁷³ Thus, effective disclosure systems become “doubly embedded.”⁵⁷⁴ The way disclosure policies affect discloser behavior is intuitive—by affecting their profits, market share, and/reputation.⁵⁷⁵ Disclosers may change their behavior, in fact, simply in anticipation that releasing information may affect one of these three things.

Thus, agencies may choose two very different courses: massive raw data dumps intended for sophisticated intermediaries, or highly distilled presentations intended for lay users. The correct choice, if one must be made, depends very much on the data and what the agency hopes to achieve by publishing it.

On one hand, simplified ratings or grades are able to distill dozens or even hundreds of different complex criteria into a single understandable metric, like restaurant hygiene grades, hospital star ratings, or five-star crash safety ratings, which are based on complex engineering standards and test results.⁵⁷⁶ Ratings and grades also combat the problem of overdisclosure. Scholars that have evaluated the effectiveness of mandatory disclosure regimes sometimes observe that parties can “overdisclose” information to try to “overwhelm and distract” the intended audience.⁵⁷⁷ Ratings can ameliorate this problem.

On the other hand, sometimes efforts to make the complex more understandable fail, as evidenced by the vague five-color scheme for communicating the threat of a terrorist attack.⁵⁷⁸ Unlike dirty restaurants or unsafe cars, it is hard for most people to understand the significance of the terror threat

572. FUNG ET AL., *supra* note 15, at 50–51.

573. *Id.* at 65–74.

574. *Id.* at 65.

575. *Id.* at 66.

576. *Id.* at 59–61.

577. Ben-Shahar & Schneider, *supra* note 17, at 700; Willis, *supra* note 231, at 790.

578. FUNG ET AL., *supra* note 15, at 61.

changing from yellow to orange, and more importantly, how to act on that signal.⁵⁷⁹ Thus, not all data is so easily distilled.

Given these considerations, should agencies design databases to be accessible to the lay public, or to be used by more sophisticated information intermediaries? An ideal answer is “both,” of course. To maximize accessibility, the data should be available in multiple formats, as many agencies now recognize, and as Data.gov demonstrates. If “both” is not a feasible option, the agency must decide whether “big data” or “small data” better achieve regulatory ends, including the relative costs of both approaches.

3. Intermediaries and Collaborative Data

Designing databases for use by third-party information intermediaries is compelling for several reasons.⁵⁸⁰ Publicly-minded watchdogs like Pro Publica, the Sunlight Foundation, and the Project on Government Oversight (POGO) can serve a translational role, sifting large amounts of data into more understandable bits.⁵⁸¹ Although these organizations focus on government transparency, they also can (and do in fact) help extract and translate information about regulated parties.⁵⁸² Even complex datasets that are not translated by agencies into ratings, grades, or other digestible metrics can be translated by thoughtful intermediaries. For example, various consumer groups have tried to translate toxic release data into more user-friendly websites.⁵⁸³ Thus, even raw data sets can be

579. *Id.*

580. For an early discussion of the use of data intermediaries, see Sage, *supra* note 5, at 1737–41.

581. *See, e.g.*, Shkabatur, *supra* note 55, at 118.

582. *See, e.g.*, Charles Ornstein et al., *Dollars for Docs*, PRO PUBLICA, <https://projects.propublica.org/docdollars/> (last visited June 6, 2017) [<https://perma.cc/W7UC-SSBF>]. The Dollars for Docs project at Pro Publica posts a searchable database of physicians that have received money from pharmaceutical and device firms, using data reported under federal law, including the Physician Payment Sunshine Act. CMS released the data, but Pro Publica gathers it in a single searchable database with rankings and analysis. *See* Charles Ornstein & Ryann Grochowski Jones, *About the Dollars for Docs Data*, PRO PUBLICA (July 1, 2015), <https://www.propublica.org/article/about-the-dollars-for-docs-data> [<https://perma.cc/6TYP-8LHQ>].

583. FUNG ET AL., *supra* note 15, at 62; *The Pollution Information Site*, SCORECARD, <http://scorecard.goodguide.com> (last visited July 16, 2017) [<http://perma.cc/B7B8-7375>]; *The Right-to-Know Network*, HOUS. CHRON., <http://www.rtknet.org> (last visited July 16, 2017) [<http://perma.cc/VL8P-QJBT>].

repurposed for lay users like consumers and other regulatory beneficiaries. Indeed, Fung, Graham, and Weil found that the most successful disclosure regimes “featured strong groups representing information users, offered benefits to at least some information disclosers, and provided comprehensible content.”⁵⁸⁴

However, translating voluminous, complex government data requires not only minimum technical and programming expertise, but also a basic understanding of the agency and its regulatory framework (and perhaps also an understanding of the regulated industry). The number of organizations that can fit comfortably into such a Venn diagram might be quite small. Indeed, even proponents of publishing raw government data concede that it is not immediately accessible to most lay users.⁵⁸⁵ And some doubt that these organizations derive their value from information supplied by the government, rather than from their own surveys and information-collecting activities.⁵⁸⁶ Still, the fact that there are individuals like Joshua Tauberer (who created Govtrack.us in his spare time)⁵⁸⁷ and Carl Malamud (who painstakingly made SEC data available online)⁵⁸⁸ demonstrate that the barriers are far from insurmountable. Moreover, these extraordinary individual efforts inspired the government to publish the data in open formats.⁵⁸⁹

Unsurprisingly, agencies are also discovering that data is becoming a more collaborative endeavor. There is optimism that once raw data is published, the private, public, and nonprofit sectors will make the data more accessible and useful to their constituents.⁵⁹⁰ Fung and colleagues note what happened with the EPA’s Toxic Release Inventory, as consumer groups like Scorecard and RTK refined the data and made it more user friendly, while the Chemical Manufacturers Association launched its own site emphasizing not only

584. FUNG ET AL., *supra* note 15, at xiv.

585. Robinson et al., *supra* note 518, at 173.

586. Ben-Shahar & Schneider, *supra* note 17, at 731–32.

587. Robinson et al., *supra* note 518, at 165–66, 171.

588. *Id.* at 166, 171.

589. *Id.* at 171.

590. FUNG ET AL., *supra* note 15, at 153.

companies' improving safety data, but also the number of jobs they created and taxes they paid by ZIP code.⁵⁹¹

As such, we might be experiencing a major shift in the government's informational role from controller to facilitator.⁵⁹² Indeed, modern agencies may be best suited to facilitating rather than controlling information—that is, gathering and publishing data, ensuring its quality, and then enabling the private and nonprofit sectors to maximize its uses.⁵⁹³ Data users might also become contributors, as in the case of consumers who report food poisoning from restaurants and thus supplement relatively infrequent restaurant inspections.⁵⁹⁴ Just as the CFPB endeavors, the government can serve as an aggregator of disparate data sources.

Craswell calls this “government-aided disclosures (GADs),” in which the government creates a baseline for disclosure, but allows companies to use the information dynamically or go beyond the baseline in some way.⁵⁹⁵ Such disclosures are mandated by government but are also integrated by disclosers because the information is useful to consumers or users.⁵⁹⁶

Another factor that improves success is whether the information varies between disclosers, such that disclosers have an incentive to race to the top.⁵⁹⁷ Cigarette brands have little incentive to highlight the Surgeon General's mandatory warnings, because the same preset warnings rotate among all products regardless of manufacturer, but they may have more incentive to reduce the tar and nicotine content of their products, which varies from brand to brand.⁵⁹⁸ Thus, data can serve as an important differentiator between competitors, and they might spend their own resources publicizing differences.⁵⁹⁹ Again, if databases aspire to affect underlying behavior and achieve regulatory ends, this is one way to do so.

In some notable instances, the federal government devotes remarkable resources to encourage users to collect and deploy

591. *Id.* at 158–60.

592. Indeed, Sage long ago called attention to the government's role as a data facilitator. Sage, *supra* note 5, at 1712.

593. See, e.g., FUNG ET AL., *supra* note 15, at 166.

594. *Id.*

595. Craswell, *supra* note 18, at 369.

596. *Id.* at 369–70.

597. *Id.* at 371.

598. *Id.*

599. *Id.* at 370–71.

certain data, as in the case of electronic health records (EHRs). Through various pieces of legislation,⁶⁰⁰ Congress not only established standards for collecting and using electronic health records, but also devoted billions in incentives—an average of more than \$40,000 per physician.⁶⁰¹

Extending the principle even further, the federal site HealthData.gov aggregates over 2,000 unique datasets from agencies like the CDC, CMS, FDA, and numerous state and local governments.⁶⁰² The goal is to put open, machine-readable data in the hands of programmers, entrepreneurs, journalists, providers, scientists, consumers, and other policymakers who might, in turn, help improve health care in the United States.⁶⁰³ Thus, given the current fascination with “big data,” it helps to remember that countless government agencies (and Congress) are helping to make data “bigger,” consonant with their traditional goal of providing public goods.⁶⁰⁴

The next generation of disclosure will thus be more collaborative, in the sense that various sectors will both contribute to and use the data.⁶⁰⁵ Indeed, there seems to be wide agreement that the government should not have a monopoly on generating data,⁶⁰⁶ but can play an important centripetal role in compiling data and helping assure their quality.

CONCLUSION

Agency databases have proliferated on the belief that markets, regulation, and democracy all thrive on transparency—that sunlight is the best disinfectant. Ideally,

600. Most notably the Health Information Technology for Economic and Clinical Health (HITECH) Act, Pub. L. No. 111-5, 123 Stat. 226 (2009) (codified as amended in scattered sections of 42 U.S.C.).

601. Madison, *supra* note 30, at 1618; David Blumenthal & Marilyn Tavenner, *The “Meaningful Use” Regulation for Electronic Health Records*, 363 NEW ENG. J. MED. 501, 501 (2010).

602. *Datasets*, HEALTHDATA.GOV, <http://www.healthdata.gov/dataset?f%5B0%5D=type%3Adataset> (last visited July 15, 2017) [<http://perma.cc/G936-V2TJ>] (sorted by “Publisher”).

603. *About*, HEALTHDATA.GOV, <http://www.healthdata.gov/content/about> (last visited July 15, 2017) [<http://perma.cc/B629-5EXX>].

604. Madison, *supra* note 30, at 1620–21; Sage, *supra* note 5, at 1771 (discussing the public good nature of information in the health industry).

605. FUNG ET AL., *supra* note 15, at 153.

606. *See, e.g.*, Esty, *supra* note 4, at 199–200.

shining a light on things like campaign contributions, pollution, and hospital outcomes will encourage more optimal behavior—or at least deter the worst of it. An added benefit is that “regulation by database” avoids the costs and formalities of traditional regulation.

But as transparency has moved online—becoming more pervasive, more powerful, and more burdened with regulatory dimensions—we also must recognize that sunlight can also blind or even burn. The case studies demonstrate how problems with accuracy, fairness, and efficacy can undermine even well-established, well-meaning data regimes. These problems can be avoided if agencies act less like passive publishers or repositories for data, and more like data stewards actively tending to a valuable (and dynamic) public good.

Policymakers must embrace the government’s role as a data steward, a sentinel that helps maximize the quality and reliability of data inputs and outputs via administrative safeguards. Thinking carefully about publication procedures, how to balance the interests of both subjects and users, how to present the data accurately and fairly, and how to maximize uses by audiences of varying sophistication can be just as resource-intensive as traditional regulation. But these steps are necessary for data to achieve “regulatory” ends. The more reliable government data are, the more they can enlighten us and deter unwanted behavior.