

REGULATION BY DATABASE

Nathan Cortez*

The federal government currently publishes 195,245 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, then, evaluates the aspirations and burdens of regulation by database. Based on case studies of six important data sets (published by the CFPB, CPSC, EPA, 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 also 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.

* Callejo Endowed Professor of Law, Gerald J. Ford Research Fellow, and Associate Dean for Research, SMU Dedman School of Law. J.D. Stanford, B.A. University of Pennsylvania. I am grateful to Nick Bagley, Glenn Cohen, James Coleman, Mark Fenster, Chris Jenks, Jeff Kahn, Kristin Madison, Frank Pasquale, Bill Sage, and to participants at Harvard Law School’s Health Law Workshop, the University of Texas Law School, and the SMU Law Faculty Forum for commenting on earlier drafts. I also thank Elise LeGros and Victoria Vish for 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 (Jun. 21, 2016); ACUS, Agency Information Dissemination in the Internet Era, <https://www.acus.gov/research-projects/agency-information-dissemination-internet-era>.

INTRODUCTION	3
I. FROM OPEN GOVERNMENT TO OPEN INDUSTRY	8
II. DATA AND ITS ASPIRATIONS.....	16
A. Market Ideals.....	17
B. Regulatory Ideals.....	18
C. Democratic Ideals.....	21
III. DATA AND ITS DISCONTENTS	22
IV. AGENCY DATABASES	29
A. The FEC’s Campaign Finance Data	29
B. The EPA’s Toxic Release Data	33
C. The CFPB’s Consumer Complaint Data	37
D. The CPSC’s Product Safety Data.....	40
E. Medicare’s Quality Data.....	43
F. The FDA’s Adverse Event Data	50
V. DESIGNING FOR OPTIMAL DISCLOSURE	54
A. Gathering and Processing the Data.....	55
1. Pre-Publication Procedures	56
2. Post-Publication Procedures for Contested Data	57
3. Considering Costs	60
4. Administrative Law Dimensions	61
B. Characterizing the Data	62
1. Identifying Sources.....	62
2. Explaining Context, Limitations	63
C. Presenting the Data.....	64
1. Raw or Polished?	64
2. Big Data or Small Data?	66
3. Intermediaries and Collaborative Data	69
CONCLUSION.....	72

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 195,245 unique data sets—roughly 150,000 from the federal government, with the remainder from sub-federal and nongovernmental sources.³ Some consequential portion of these includes 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 to pursue “regulatory” aims—to change 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

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 (2014).

³ Data.gov, Search for a Dataset, <https://catalog.data.gov/dataset> (last visited Feb. 7, 2017) (see “Organization Types” in the left sidebar). Data.gov currently includes data from 80 federal agencies and subagencies. See Data.gov, Federal Agency Participation, <http://www.data.gov/metrics> (last visited Feb. 7, 2017).

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).

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 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 regulatory method internationally.¹⁵

Though 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 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 now much less static (aimed narrowly at helping potential users of disclosed information make better decisions) and more dynamic (aimed more broadly at trying to change the disclosers' underlying behavior).¹⁸ Frequently, the real party being targeted by mandatory disclosure is not the consumer, but the discloser,¹⁹

⁶ Nutritional Labeling Education Act of 1990, _____.

⁷ U.S. Department of Energy, Download Fuel Economy Data, Fueleconomy.gov, <http://www.fueleconomy.gov/feg/download.shtml> (providing data back to 1978).

⁸ See Part II.E, *infra*.

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

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 Paredes, *Blinded by the Light: Information Overload and Its Consequences for Securities Regulation*, 81 WASH. U. L. Q. 417 (2003).

¹¹ Sex Offender Registration and Notification Act, Pub. L. No. __ (codified at ____).

¹² Transportation Recall Enhancement, Accountability, and Documentation (TREAD) Act, Pub. L. No. 106-414, 114 Stat. 1800 (*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, Pub. L. No. _____.

¹⁴ Occupational Safety and Health Administration, Hazard Communication Standard, 29 C.F.R. § 1910.1200 (requiring material safety data sheets in the workplace).

¹⁵ FUNG ET AL., *supra* note __, at 127-50; 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 Ben-Shahar & Schneider, *supra* note __, at 744.

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 __, at 744.

under the Brandeisian logic that shining a light on undesired behavior will deter it, or at least make it more costly.

Disclosure, then, remains in bloom. 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 web sites FOIA.gov and Data.gov. It is possible, if not likely, that online publication of data 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.

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 web sites, mobile applications, and other informational products that rely on government data. Perhaps the clearest 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 innovation that incorporates geospatial location data,²² such as navigation, restaurant, and social media apps. Indeed, when President Obama announced his “Open Data Policy,” he remarked that “This kind of innovation and ingenuity has the potential to transform the way we do almost everything.”²³ Lest the reader discount this as hyperbole, disclosure policies are sometimes burdened by their virtually limitless ambitions—often being justified as promoting “autonomy, dignity, civility, community, citizenship, economic growth, and a variety of other

²⁰ Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government (Jan. 21, 2009), 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.COM (Dec. 15, 2016), at <http://fivethirtyeight.com/features/how-trumps-white-house-could-mess-with-government-data/>.

²² SCOTT PACE ET AL., *THE GLOBAL POSITIONING SYSTEM: ASSESSING NATIONAL POLICIES*, Appendix B: GPS History, Chronology, and Budgets (Rand Corp. 1995), https://www.cs.cmu.edu/~sensing-sensors/readings/GPS_History-MR614.appb.pdf; HealthData.gov, About, at <http://www.healthdata.gov/content/about>.

²³ White House, *President Obama Speaks on Innovation and Manufacturing*, (May 9, 2013), <https://www.whitehouse.gov/photos-and-video/video/2013/05/09/president-obama-speaks-innovation-and-manufacturing>; see also Memorandum for the Heads of Executive Departments and Agencies from Sylvia M. Burwell et al., *Open Data Policy—Managing Information as an Asset* (May 9, 2013), <https://www.whitehouse.gov/sites/default/files/omb/memoranda/2013/m-13-13.pdf>.

virtues.”²⁴ Part II examines these aspirations in light of emerging evidence.

Given these policy justifications, then, many rightly wonder who could possibly oppose providing more information to the public.²⁵ But as ubiquitous as disclosure has become, criticisms have emerged from scholars who doubt that it is “an unalloyed good.”²⁶ I consider the burdens of “regulation by database” in detail in Part III. 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.²⁷

As Schauer notes, policymakers rarely question whether the burdens of disclosure outweigh its purported benefits.²⁸ One goal of this Article, then, 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” relating to the use of 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

²⁴ Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 734 (2011).

²⁵ FUNG ET AL., *supra* note __, 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.” Fenster, *supra* note __, at 888-89.

²⁶ See, e.g., Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1342; Ben-Shahar & Schneider, *supra* note __; 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 & SOCIETY 1 (Jan.-Jun. 2016), at <http://journals.sagepub.com/doi/pdf/10.1177/2053951715621568>.

²⁷ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1342.

²⁸ Ben-Shahar & Schneider, *supra* note __, at 683 (“[L]awmakers rarely inquire into the effectiveness or burden of disclosure.”).

		chemicals by facility name, address.
FDA	FDA Adverse Event Reporting System (FAERS), Manufacturer and User Facility Device Experience (MAUDE)	Adverse events 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 change 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 probably 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. This includes recommending 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 accurately, and drawing lessons for agencies considering whether to publish “raw” or “polished” data, as well as “big” or “small” data.

I envision a decidedly modern role for the government as a “data steward”²⁹ rather than merely as a data source or publisher. For example, there are smart ways that the government can help gather and generate data—making data “bigger.”³⁰ But for some data, it might be preferable for the government to distill the data and make it more user-friendly—making data “smaller.”³¹ 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 is communicated clearly to users. 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.³² We might refer to

²⁹ 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).

³⁰ Madison, *supra* note __, at 1627-28.

³¹ Madison, *supra* note __, at 1627-28.

³² ACUS, *Adoption of Recommendations*, 81 Fed. Reg. 40,259 (Jun. 21, 2016); ACUS, *Agency Information Dissemination in the Internet Era*, <https://www.acus.gov/research-projects/agency-information-dissemination-internet-era>;

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 alter the underlying conduct of disclosers. Failed data policies, conversely, can produce incomplete or “gerrymandered” information, create a false sense of security, waste resources, undermine public trust, and even put lives at risk.³³ To design databases for optimal disclosure, we must also appreciate that disclosure is an exercise of power. Schauer calls it “transparency as regulation,” recognizing that “for one person or institution to have information about another is for the former to have power over the latter.”³⁴ Similarly, “transparency” can be recharacterized as “adverse transparency” if the information is unflattering or harmful to the subject in some way.³⁵ How, then, can policymakers exercise this power responsibly and legitimately? If we value disclosure and accept it as a baseline, how do we manage it? As federal data is used more and more to achieve regulatory ends, both the means and ends require more critical examination.

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,³⁶ 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,³⁷ sitting comfortably “among the pantheon of great political virtues.”³⁸

But today’s transparency arrived only after decades of reforms. The New Deal birthed a generation of new executive agencies and corresponding regulations. But agencies published their regulations at will, if at all. One of the earliest

³³ FUNG ET AL., *supra* note __, at 172.

³⁴ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1347.

³⁵ Sarah Taylor Roller, Raqiyyah R. Pippins, & Jennifer W. Ngai, *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).

³⁶ 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,” U.S. CONST. ART. I § 5, which some view as a deliberate departure from the secrecy practiced by the British Parliament. James Brudney, *Canon Shortfalls and the Virtues of Political Branch Interpretive Assets*, 98 *CAL. L. REV.* 1199, 1220 (2010).

³⁷ William Funk, *Public Participation and Transparency in Administrative Law—Three Examples as an Object Lesson*, 61 *ADMIN. L. REV.* 171 (2009).

³⁸ Mark Fenster, *The Opacity of Transparency*, 91 *IOWA L. REV.* 885, 888 (2006).

efforts to address the lack of transparency among federal agencies was the Federal Register Act of 1935,³⁹ 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.”⁴⁰

The Act was motivated in part by the famous “hot oil” case, *Panama Refining*.⁴¹ The “hot oil” law was part of the National Industrial Recovery Act, the flagship New Deal bill passed during the Great Depression.⁴² The 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. 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.”⁴³ Just weeks before oral argument in the case, Erwin Griswold published a law review article, *Government in Ignorance of the Law*, arguing for a Federal Register-like system to publish executive branch laws.⁴⁴ The “furor” over the case⁴⁵ 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.”⁴⁶ 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.”⁴⁷

³⁹ Pub. L. No. 74-220, ch. 417, 49 Stat. 500 (1935).

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

⁴¹ 293 U.S. 388 (1935).

⁴² Pub. L. No. 73-67, 48 Stat. 195 (1933).

⁴³ Peter Strauss et al., *Gellhorn and Byse’s Administrative Law: Cases and Comments* 446 (11th ed. 2011).

⁴⁴ Erwin N. Griswold, *Government in Ignorance of the Law—A Plea for Better Publication of Executive Legislation*, 48 HARV. L. REV. 198 (1934). Strauss et al. explain that 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 __, at 446 n.4.

⁴⁵ *Id.* (citing Mary Whisner, *Practicing Reference ...: 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)).

⁴⁶ *Cervase v. Office of the Federal Register*, 580 F.2d 1166, __ (3d Cir. 1978) (“The basic object of this statutory reform was to eliminate secret law.”).

⁴⁷ Office of the Federal Register, *The Federal Register: March 14, 1936 – March 14, 2006: A Brief History Commemorating the 70th Anniversary of the Publication of the First Issue of the Federal Register* 2, https://www.federalregister.gov/uploads/2011/01/fr_history.pdf; STRAUSS ET AL., *supra* note __, at 445-46.

Thus, before 1935, agencies did not even have to publish the regulations they imposed.⁴⁸

Publication requirements were further embedded in American law by the Administrative Procedure Act of 1946 (APA).⁴⁹ The APA was a response, in part, to complaints from industry that administrative agencies during the New Deal were too opaque and insular, particularly towards the private interests most affected by regulation.⁵⁰ 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.⁵¹ New APA procedures creating “notice and comment” rulemaking have since become a hallmark of citizen participation in government.⁵² 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.⁵³

However, many eventually came to view APA section 3 more as a tool to *withhold* information than disclose it. 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. Such documents were only accessible, just as before 1935, on a haphazard basis.⁵⁴

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.”⁵⁵ 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.⁵⁶ As Bill Funk observes, the APA addresses participation in

⁴⁸ Funk, *supra* note __, at 173. 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).

⁴⁹ Pub. L. No. 79-404, 60 Stat. 237 (1946) (codified as amended in scattered sections of 5 U.S.C.).

⁵⁰ William Funk, Public Participation and Transparency in Administrative Law – Three Examples as an Object Lesson, 61 ADMIN. L. REV. 171, 172-73, 178 (2009).

⁵¹ APA § 3, 60 Stat. 237 (1946).

⁵² 5 U.S.C. § 553; Shkabatur, *supra* note __, at 85; Beth Simone Noveck, *The Electronic Revolution in Rulemaking*, 53 EMORY L.J. 433, 517 (2004).

⁵³ 5 U.S.C. § 552(a)(1).

⁵⁴ STRAUSS ET AL., *supra* note __, at 451.

⁵⁵ Funk, *supra* note __, at 173.

⁵⁶ Funk, *supra* note __, at 173; FINAL REPORT OF THE ATTORNEY GENERAL’S COMMITTEE ON ADMINISTRATIVE PROCEDURE 2 (1941), available at

rulemaking to “interested persons,” meaning those with a “direct and palpable interest,”⁵⁷ and required that public records be made available to persons “properly and directly concerned,” rather than the general public.⁵⁸ Thus, the APA’s disclosure provisions were aimed to inform regulated parties, not shine a light on them, as became the focus decades later.

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.⁵⁹ The modern open government movement really began in 1967, when Congress passed the Freedom of Information Act (FOIA),⁶⁰ 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.⁶¹ President Lyndon Johnson, who signed the bill reluctantly and only under pressure from the press corps,⁶² 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.”⁶³ Half a century later, modern scholars still acknowledge FOIA’s importance to our democratic government.⁶⁴ In requiring the government to make its records available upon request, unless specifically exempted, FOIA created a “strong presumption in favor of disclosure.”⁶⁵ 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.⁶⁶

<http://www.law.fsu.edu/library/admin/1941report.html>.

⁵⁷ Funk, *supra* note __, at 174.

⁵⁸ Administrative Procedure Act, ch. 324, § 3(c), 60 Stat. 237 (1946) (current version at 5 U.S.C. § 552(a)(3)). Funk explains that these goals derive in part from the canonical work by James Landis, *The Administrative Process*. Funk, *supra* note __, at 177 (citing JAMES M. LANDIS, *THE ADMINISTRATIVE PROCESS* (1938)).

⁵⁹ Funk, *supra* note __, at 178-80.

⁶⁰ Pub. L. No 90-23, 81 Stat. 54 (1967). 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, *supra*, which was identical in substance. Vladeck has characterized FOIA as “truly an experiment in open government.” Vladeck, *supra* note __, at 1795.

⁶¹ See, e.g., 112 CONG. REC. H13007 (daily ed. Jun. 20, 1966) (statement of Rep. John Moss).

⁶² Vladeck, *supra* note __, at 1798.

⁶³ Presidential Statement on Signing the Freedom of Information Act, S.1160 __ WEEKLY COMP. PRES. DOC. __ (Jul. 4, 1966), available at <http://nsarchive.gwu.edu/NSAEBB/NSAEBB194/Document%2031.pdf>.

⁶⁴ David C. Vladeck, *Information Access – Surveying the Current Legal Landscape of Federal Right-to-Know Laws*, 86 TEX. L. REV. 1787 (2008) (noting that FOIA “embodies the ideal that information is the lifeblood of democracy.”).

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

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

FOIA, of course, has been criticized for falling short of its lofty goals of pursuing democracy through transparency and accountability.⁶⁷ A major complaint is that FOIA produces transparency only by request. 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.”⁶⁸ Scholars have also criticized FOIA for being too malleable, particularly by administrations that construe its disclosure requirements narrowly and exemptions broadly (the George W. Bush administration is frequently cited).⁶⁹ As such, FOIA envisions passive disclosure by agencies rather than actively disclosing or even publicizing the information they hold.⁷⁰ Moreover, the basic premise of FOIA’s “request-and-wait-for-a-response approach” is seen as obsolete in the Internet era.⁷¹

Several laws followed FOIA. In the 1990s, as the Internet came of age, a series of laws pushed the government to use it. For example, the Paperwork Reduction Act of 1995,⁷² the Electronic Freedom of Information Act of 1996 (sometimes called “E-FOIA”),⁷³ and the Government Paperwork Elimination Act of 1998⁷⁴ directed federal agencies to use the Internet to publish more information online and to “improve the productivity, efficiency, and effectiveness of Federal programs.”⁷⁵ In particular, E-FOIA required agencies to publish online their final opinions and orders, as well as records likely to be requested.⁷⁶ Thus spawned the brief surge in agency online “reading rooms.”

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

⁶⁷ Shkabatur, *supra* note __, at 88-89 (citing several articles and at least one law, the OPEN Government Act of 2007, Pub. L. No. 110-175, 121 Stat. 2524 (codified in scattered sections of 5 U.S.C.)).

⁶⁸ 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 __, at 89; Vladeck, *supra* note __, at 1789.

⁶⁹ Scholars often point the George W. Bush administration on this point. 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 __, at 1790; Shkabatur, *supra* note __, at __.

⁷⁰ Cortez, *supra* note __, at __.

⁷¹ Vladeck, *supra* note __, at 1792-93.

⁷² Pub. L. No. 104-13, § 2, 109 Stat. 163, 167 (codified at 44 U.S.C. § 3504(h)(5)).

⁷³ Pub. L. No. 101-231, 110 Stat. 347 (codified at 5 U.S.C. § 552).

⁷⁴ Pub. L. No. 105-277, div. C, tit. XVII, § 1702, 112 Stat. 2681 (codified at 44 U.S.C. § 3504).

⁷⁵ Pub. L. No. 104-13, § 2, 109 Stat. 163, 167 (codified at 44 U.S.C. § 3504(h)(5)).

⁷⁶ 5 U.S.C. § 552(a)(2)(D).

information not just about their own activities, but about regulated parties as well.

In the 2000s, as federal agencies built sprawling web sites,⁷⁷ 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.⁷⁸ The Act also created the Office of Electronic Government within OMB, headed by a Chief Information Officer that would coordinate with agencies through a Council.⁷⁹ 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, this time using Internet technologies.⁸⁰ 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.⁸¹

Likewise, during this decade Congress also passed laws requiring more transparency in federal spending, directing the OMB to publish online the details of federal grants, loans, and contracts.⁸² Today, the public can search the federal web site USAspending.gov to view entities that have received federal money,⁸³ or more specific sites like Recovery.gov to see how the federal government has spent money from the economic stimulus package of 2009.⁸⁴ 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.⁸⁵ As Vladeck

⁷⁷ James O'Reilly, *Libels on Government Websites: Exploring Remedies for Federal Internet Defamation*, 55 ADMIN. L. REV. 507 (2003).

⁷⁸ Pub. L. No. 107-347, 116 Stat. 2899 (codified as amended in scattered sections of the U.S. Code).

⁷⁹ *Id.*

⁸⁰ *Id.* at § 101(a), 116 Stat. 2902 (defining “electronic government”).

⁸¹ Open Government Act of 2007, Pub. L. No. 110-175, 121 Stat. 2524 (codified in scattered sections of 5 U.S.C. § 552).

⁸² E-Government Act of 2002, Pub. L. No. 107-347, 116 Stat. 2899; Federal Funding Accountability and Transparency Act of 2006, Pub. L. No. 109-282, 120 Stat. 1186 (2006) (codified at 31 U.S.C. § 6101).

⁸³ See USAspending.gov, at <http://www.usaspending.gov>.

⁸⁴ Recovery.gov, U.S. Recovery Accountability and Transparency Board, at <http://www.recovery.gov>. 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 Act (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, at <http://thomas.loc.gov/cgi-bin/bdquery/z?d112:h.r.02146>.

⁸⁵ Pub. L. No. 109-282 § 2(b)(1), 120 Stat. at 1187; see also USAspending.gov at

notes, the site “was able to piggyback on the work of OMB Watch [now the Center for Effective Government], a nonprofit watchdog organization that with foundation support had already constructed a comprehensive, searchable database that is also available free of charge to the public.”⁸⁶ 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).⁸⁷ 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.”⁸⁸ 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.⁸⁹ In 2002, the OMB finalized guidelines implementing the Act,⁹⁰ followed by agencies issuing their own guidelines.⁹¹

These IQA guidelines would seem to be well situated 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.”⁹² And the OMB guidelines define “information” as “any communication or representation of knowledge such as facts or data, in any medium or form,”⁹³ including “information that an agency disseminates from a web page.”⁹⁴ 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.”⁹⁵ The

<http://www.usaspending.gov>.

⁸⁶ Vladeck, *supra* note __, at 1829-30 (citing About FedSpending.org, at <http://www.fedspending.org/aboutthissite.php>).

⁸⁷ 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.

⁸⁸ *Id.* 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.

⁸⁹ *Id.*

⁹⁰ 66 Fed. Reg. 34489 (Jun. 28, 2001) (proposed guidelines); 66 Fed. Reg. 49718 (Sep. 28, 2001); 67 Fed. Reg. 369 (Jan. 3, 2002); 67 Fed. Reg. 8452 (Feb. 22, 2002).

⁹¹ OMB, Agency Information Quality Guidelines, https://www.whitehouse.gov/omb/inforeg_agency_info_quality_links/ (last visited June 15, 2015).

⁹² 44 U.S.C. § 3504(d)(1).

⁹³ 67 Fed. Reg. at 8458-59.

⁹⁴ *Id.* at 8460.

⁹⁵ 67 Fed. Reg. at 377.

guidelines also exempt “adjudicative processes.”⁹⁶ These exemptions might thus exclude important agency databases, such as the CFPB’s Consumer Complaint Database (discussed in Part IV.C, *infra*).

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,⁹⁷ as well as a memorandum on FOIA.⁹⁸ 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.⁹⁹

In December 2009, the OMB published the Open Government Directive,¹⁰⁰ 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.”¹⁰¹ 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 had previously not been available.¹⁰² Within 60 days, each agency was to create an Open Government web page. Today, 80 different federal agencies and subagencies have posted 195,245 datasets on Data.gov.¹⁰³

In 2011, the Justice Department created FOIA.gov, a web site that publicizes data on how agencies have performed their FOIA duties.¹⁰⁴ 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

⁹⁶ *Id.*

⁹⁷ 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. Jennifer Shkabatur, *Transparency With(out) Accountability: Open Government in the United States*, 31 YALE L. & POL’Y REV. 79, 80 (citing Open Government Partnership, <http://www.opengovpartnership.org>).

⁹⁸ Memorandum for the Heads of Executive Departments and Agencies: Freedom of Information Act, 74 Fed. Reg. 4683 (Jan. 21, 2009).

⁹⁹ Gellhorn & Byse’s *Administrative Law*, at 441. Of course, even the Obama Administration has been criticized for rejecting more transparency in matters of terrorism and national security. Jeff Kahn, [TITLE], in *CAMBRIDGE HANDBOOK OF SURVEILLANCE LAW* [DATE].

¹⁰⁰ Memorandum from Peter R. Orszag, Dir., Office of Mgmt. & Budget, to the Heads of Exec. Dep’ts and Agencies (Dec. 8, 2009).

¹⁰¹ *Id.*

¹⁰² *Id.* at 7.

¹⁰³ Federal Agency Participation, Data.gov, at <http://www.data.gov/metric> (last visited Feb. 9, 2017).

¹⁰⁴ FOIA.gov, U.S. Department of Justice, at <http://www.foia.gov>.

shaming” to encourage agencies to be more responsive to FOIA requests.¹⁰⁵ Still, scholars question how effectively “naming and shaming” is at convincing underperforming agencies to increase their responsiveness to FOIA requests.¹⁰⁶

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.¹⁰⁷ Thus, the current gestalt that pursues transparency from industry owes much to earlier right-to-know efforts pursuing transparency from government.¹⁰⁸ 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.

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.”¹⁰⁹ Contemporary scholarship, of course, has focused on the many ways in which Internet technologies have facilitated communication between the government and the public.¹¹⁰ But after years of scholarly praise of the Internet’s role in facilitating transparency, accountability, and democracy, inevitable critiques have emerged.¹¹¹ Still, disclosure frequently is offered as a tool that can achieve market, regulatory, and democratic ideals.¹¹²

¹⁰⁵ Shkabatur, *supra* note __, at 100.

¹⁰⁶ Shkabatur, *supra* note __, at 100.

¹⁰⁷ Also making this observation are FUNG ET AL., *supra* note __, at xii-xiii. They call this “targeted transparency.”

¹⁰⁸ FUNG ET AL., *supra* note __, 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.

¹⁰⁹ Ben-Shahar & Schneider, *supra* note __, at 734.

¹¹⁰ For just a small sample, *see, e.g.*, Bruce Bimber, *Information and American Democracy: Technology and the Evolution of Political Power* (2003); Stephen Coleman & Jay G. Blumler, *The Internet and Democratic Citizenship* (2009).

¹¹¹ *See, e.g.*, MATTHEW HINDMAN, *THE MYTH OF DIGITAL DEMOCRACY* (2003); EVGENY MOROZOV, *THE NET DELUSION* (2012).

¹¹² Indeed, Sage evaluated these types of justifications almost 20 years ago, characterizing them somewhat differently as four separate rationales (competition, agency, performance, and democratic). He also observes, however, that disclosure rationales can often be contradictory or at

A. Market Ideals

A frequently-invoked rationale for regulatory disclosures is that disclosure can improve consumer decisionmaking, facilitate markets, and “protect the naïve from the sophisticated.”¹¹³ 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 undermine the market.¹¹⁴ 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.”¹¹⁵

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.¹¹⁶ 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.¹¹⁷ The Bureau concludes that after the data is made public, “The marketplace of ideas then does the rest.”¹¹⁸

Disclosure thus satisfies both our free-market intuitions¹¹⁹ 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 services¹²⁰—a refrain

least in tension. See, Sage, *Regulating Through Disclosure*, *supra* note __.

¹¹³ Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 649 (2011).

¹¹⁴ George A. Akerloff, *The Market for “Lemons”: Quality Uncertainty and the Market Mechanism*, 84 Q.J. ECON. 488 (1970) (using as an example the used car market).

¹¹⁵ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1350.

¹¹⁶ CFPB, *Disclosure of Certain Credit Card Complaint Data*, 76 Fed. Reg. 76,628, 76,630 n.9 (Dec. 8, 2011).

¹¹⁷ 76 Fed. Reg. at 76,631.

¹¹⁸ 76 Fed. Reg. at 76,631.

¹¹⁹ Ben-Shahar & Schneider, *supra* note __, at 681.

¹²⁰ Dodd-Frank Wall Street Reform and Consumer Protection Act, Pub. L. No. 111-203, 124 Stat. 1376 (2010) (codified in scattered sections of 12 U.S.C.); 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”).

repeated frequently by the Bureau in its own publications.¹²¹ 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.”¹²²

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.”¹²³ Countless mandatory disclosure laws rely on this logic.¹²⁴ The CFPB’s complaint data, the CPSC’s product safety data, and countless other data sets are predicated on consumers using the data to vote with their wallets, avoiding sub-standard performers.

Finally, corporations and industry groups often “urge greater transparency as an alternative to allegedly more heavy-handed regulation.”¹²⁵ Of course, many scholars embrace this view as well.¹²⁶ 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.”¹²⁷ 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 and publicly accessible will think twice before acting in ways that harm investors.¹²⁸ Similarly, requiring hospitals to publish mortality rates is really a device to

¹²¹ See, e.g., 78 Fed. Reg. 21,225.

¹²² 78 Fed. Reg. at 21,220.

¹²³ Ben-Shahar & Schneider, *supra* note __, at 681.

¹²⁴ Ben-Shahar & Schneider, *supra* note __, at 681 (citing examples from the FTC and others).

¹²⁵ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1341 (citing L. Gordon Crovitz, Op-Ed., *Transparency is More Powerful Than Regulation*, WALL ST. J., Mar. 30, 2009, at A21; Richard H. Thaler & Cass R. Sunstein, Op-Ed., *Disclosure is the Best Kind of Credit Regulation*, WALL ST. J., Aug. 13, 2008, at A17).

¹²⁶ See, e.g., Crovitz, *supra* note __; Thaler & Sunstein, *supra* note __.

¹²⁷ FUNG ET AL., *supra* note __, at 5.

¹²⁸ Schauer, *Transparency in Three Dimensions*, *supra* note __, 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, *Regulating Through Information*, *supra* note __, at 1780 (noting that the “SEC penalizes faulty disclosure, not faulty performance...”).

encourage hospitals to *reduce* mortality rates.¹²⁹ Again, the same logic motivates countless disclosure regimes, with a long lineage. In 1796, Jeremy Bentham observed that “the more strictly we are watched, the better we behave.”¹³⁰

Agencies also frequently use databases to publish compliance and enforcement data. Scholars have called for agencies to actively publish enforcement records that are available under FOIA but must be requested.¹³¹ 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.”¹³² 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).¹³³

Enforcement data may also help counter underenforcement by agencies, what Matthew Stephenson calls “agency slack.”¹³⁴ Scholars note widespread underenforcement by a variety of agencies, in a variety of contexts.¹³⁵ Even when agencies do pursue regulatory violations, they often fail to enforce them.¹³⁶ Underenforcement may derive from several sources—insufficient agency

¹²⁹ Michael B. Rothberg et al., *Choosing the Best Hospital: The Limitations of Public Quality Reporting*, 27 HEALTH AFF. 1680 (2008); Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1348.

¹³⁰ JEREMY BENTHAM, *Farming Defended*, in 1 WRITINGS ON THE POOR LAWS 276, 277 (Michael Quinn ed., Oxford University Press 2001 (1796)); Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1352.

¹³¹ Vladeck, *supra* note __, at 1830-31.

¹³² Vladeck, *supra* note __, at 1830.

¹³³ Vladeck, *supra* note __, at 1830; About the Transactional Records Access Clearinghouse, <http://trac.syr.edu/aboutTRACgeneral.html>. 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. FTC, Consumer Sentinel Network, at <https://www.ftc.gov/enforcement/consumer-sentinel-network>; Cortez, ACUS Report, *supra* note __, at __.

¹³⁴ Matthew C. Stephenson, *Public Regulation of Private Enforcement*, 91 VA. L. REV. 93, 110 (2005).

¹³⁵ 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); Matthew C. Stephenson, *Public Regulation of Private Enforcement: The Case for Expanding the Role of Administrative Agencies*, 91 VA. L. REV. 93, 116-17 (2005); Rena Steinzor, *The Truth About Regulation in America*, 5 HARV. L. & POL’Y REV. 323 (2011).

¹³⁶ 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).

resources, ideology, anti-regulatory pressures, political oversight, inertia, or agency self-interest.¹³⁷ Regardless of the contributors, making compliance and enforcement data public might inspire agencies to reach more optimal levels of enforcement—or even companies to reach more optimal levels of compliance.

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).¹³⁸ The CFPB itself notes that 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.¹³⁹ 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.¹⁴⁰ Technologists have envisioned the government serving as a “platform” for innovation by providing data that inspires outside parties to create innovative uses of it.¹⁴¹ 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.¹⁴²

The Obama administration has pursued several crowdsourcing initiatives, published on web sites like Challenge.gov, which features various prize competitions sponsored by over 80 federal agencies.¹⁴³ The site claims that the government has awarded “[m]ore than \$220 million in prize money” since 2010.¹⁴⁴ The stated goal is to seek “innovative solutions from the public, bringing the best ideas and talent together to solve mission-centric problems.”¹⁴⁵

¹³⁷ Stephenson, *supra* note __, at 110.

¹³⁸ 77 Fed. Reg. at 37,562 (citing M. Beck, *Searching for Side Effects*, WALL ST. J. ONLINE (Jan. 31, 2012) (citing products by AdverseEvents, Inc. and Clarimed LLC)).

¹³⁹ Cortez, ACUS Report, *supra* note __, at 66.

¹⁴⁰ Crowdsourcing, Definition, Merriam-Webster Online Dictionary, at <http://www.merriam-webster.com/dictionary/crowdsourcing>.

¹⁴¹ 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 __, at 110.

¹⁴² Shkabatur, *supra* note __, at 109.

¹⁴³ About Challenge.gov, at <http://www.challenge.gov/about/>.

¹⁴⁴ Id.

¹⁴⁵ Id.

Indeed, even the administration’s original Open Government Directive offered contests and prizes that incentivized the public to “tinker” with the data released.¹⁴⁶ 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.¹⁴⁷ Transparency is often assumed to be a precondition for accountability.¹⁴⁸ Leading thinkers—from John Milton, John Stuart Mill, James Madison, Oliver Wendell Holmes (father of the marketplace of ideas),¹⁴⁹ and Louis Brandeis—have long drawn an explicit link between transparency, accountability, and democracy.¹⁵⁰ Centuries of writing is filled with glittering 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.¹⁵¹ Seen in this way, agency databases are a logical vehicle for both technological and democratic progress.

Government databases can also serve an expressive function by acting as a conduit for consumers to air their grievances.¹⁵² Presidents Kennedy and Nixon

¹⁴⁶ Memorandum from Jeffrey D. Zients, Deputy Dir. for Mgmt., Exec. Office of the President, to the Heads of Exec. Dep’ts and Agencies (Mar. 8, 2010); Shkabatur, *supra* note __, at 111 n.173.

¹⁴⁷ Shkabatur, *supra* note __, at 80; Adam M. Samaha, Government Secrets, Constitutional Law, and Platforms for Judicial Intervention, 53 UCLA L. REV. 909, 917 (2006).

¹⁴⁸ Open Government Memorandum, at 4685; Fenster, *supra* note __, at 894-99; Shkabatur, *supra* note __, at 83.

¹⁴⁹

¹⁵⁰ 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 __, at 24, note, this language is carved on the Library of Congress building); *Abrams v. United States*, 250 U.S. 616, 630 (1919) (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.

¹⁵¹ Schauer, Transparency in Three Dimensions, *supra* note __, 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 __, 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.*

¹⁵² 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).

both promulgated a Consumer Bill of Rights to pose the government as an intermediary or a tribunal through which consumer complaints could be given voice.¹⁵³ 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."¹⁵⁴ The CFPB's Consumer Complaint Database essentially serves this role today.

In short, there are compelling reasons why data disclosure appeals so much to policymakers. Disclosure seems consistent with free-market and autonomy principles, and seems to be an easy and effective intervention compared to more traditional regulation.¹⁵⁵ Politically, regulation by disclosure is cast as a "path of least resistance for administrative agencies seeking to promote meaningful change."¹⁵⁶ Indeed, calls for disclosure and transparency are often justified, on a more fundamental plane, as furthering the pursuit of truth, knowledge, and societal progress.¹⁵⁷

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.

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.¹⁵⁸ Although scholarship on these harms is not voluminous, the harms are relatively well documented.¹⁵⁹ 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

¹⁵³ *Id.* at 76; Special Message to the Congress on Protecting the Consumer Interest, 1962 Pub. Papers 235, 236; Special Message to the Congress on Consumer Protection, 1969 Pub. Papers 883, 883.

¹⁵⁴ Porter, *supra* note __, at 76.

¹⁵⁵ Ben-Shahar & Schneider, *supra* note __, at 681-82

¹⁵⁶ Sage, Regulating Through Disclosure, *supra* note __, at 1772.

¹⁵⁷ Schauer, Transparency in Three Dimensions, *supra* note __, at 1350.

¹⁵⁸ See, e.g., Ernest Gellhorn, *Adverse Publicity by Administrative Agencies*, 86 HARV. L. REV. 1380 (1973); Nathan Cortez, *Adverse Publicity by Administrative Agencies in the Internet Era*, 2011 BYU L. REV. 1371 (2011); Cortez, ACUS Report, *supra* note __, at 9-12.

¹⁵⁹ 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 __; Lars Noah, Administrative Arm-Twisting in the Shadow of Congressional Delegations of Authority, 1997 WIS. L. REV. 874.

million in lost sales.¹⁶⁰ And there are numerous instances in which a company's stock value plummeted after an agency announcement criticized the company.¹⁶¹ 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."¹⁶²

Questions about the value of so-called "naming and shaming" have crept into various disciplines.¹⁶³ 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 scale shame effectively to change the behavior of corporations or even governments.¹⁶⁴ Indeed, organizations like Wikileaks derive tremendous power and influence from disclosure—the kind of non-governmental power normally reserved for mainstream media.¹⁶⁵ Less sanguine views argue that "shaming is the very antithesis of the law,"¹⁶⁶ particularly when wielded by the government. For example, Donald Trump's Twitter posts, the focus of significant hand-wringing, 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.¹⁶⁷ After the episode, some investors and market analysts began to monitor Trump's tweets for potential market-moving proclamations.¹⁶⁸

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

¹⁶⁰ Denis G. Maki, *Coming to Grips with Foodborne Infection—Peanut Butter, Peppers, and Nationwide Salmonella Outbreaks*, 360 *NEW ENG. J. MED.* 949 (2009).

¹⁶¹ For examples, see Cortez, *Adverse Publicity*, *supra* note __; Cortez, ACUS Report, *supra* note __, at 9-12.

¹⁶² Vladeck, *supra* note __, at 1793 (noting, however, that secrecy should give way to publication, for example, when necessary to avoid death or serious injury).

¹⁶³ See, e.g., Brent Fisse & John Braithwaite, *The Impact of Publicity on Corporate Offenders* (1983).

¹⁶⁴ Jennifer Jacquet, *Is Shame Necessary: New Uses for an Old Tool* (2015).

¹⁶⁵ Shkabatur, *supra* note __, at 135-36.

¹⁶⁶ Eric Posner, *A Terrible Shame: Enforcing Moral Norms Without Law Is No Way to Create a Virtuous Society*, *SLATE.COM* (Apr. 9, 2015).

¹⁶⁷ Phillip Bump, *Did Donald Trump Tank Boeing's Stock Because He Was Mad about a News Article?*, *WASH. POST*, Dec. 6, 2016, at <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/>.

¹⁶⁸ Alyssa Sims, *The Stock Market Needs to Gird Itself for Trump's Tweets*, *SLATE.COM*, Dec. 16, 2016, at http://www.slate.com/articles/technology/future_tense/2016/12/the_stock_market_needs_to_gird_itself_for_trump_s_tweets.html.

decisions.¹⁶⁹ 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.”¹⁷⁰ 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.¹⁷¹ Although their bearish views are challenged,¹⁷² 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 their most useful, data that purport to present truthful, objective information must meet those standards. Otherwise, they risk succumbing to a problem known well to computer and software engineers: “Garbage in, garbage out.” 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.¹⁷³ 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.¹⁷⁴ The FEC’s campaign finance data is widely acknowledged to be incomplete.¹⁷⁵ The FDA’s medical device database, which tracks problems with devices, and is subject to mandatory reporting requirements, is undermined dramatically by under-reporting.¹⁷⁶ Even relatively non-controversial databases like USAspending.gov and Recovery.gov have been found to have widespread inaccuracies.¹⁷⁷ Both the OMB and GAO have admitted that data on USAspending.gov has been inaccurate, untimely, or incomplete.¹⁷⁸ An independent review by the Sunlight Foundation found that the site had “over 1.2 trillion dollars’ worth of

¹⁶⁹ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1351.

¹⁷⁰ Vladeck, *supra* note __, at 1792.

¹⁷¹ Ben-Shahar & Schneider, *supra* note __.

¹⁷² Craswell, *supra* note __; 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).

¹⁷³ Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683 (1999).

¹⁷⁴ 900 F. Supp. 2d 572.

¹⁷⁵ See Part II.A, *supra*.

¹⁷⁶ U.S. Government Accountability Office (GAO), *Medical Device Reporting: Improvements Needed in FDA’s System for Monitoring Problems with Approved Devices* (Jan. 29, 1997), at <http://www.gao.gov/products/HEHS-97-21>.

¹⁷⁷ See, e.g., Shkabatur, *supra* note __, at 103.

¹⁷⁸ Memorandum from Peter R. Orszag, Dir. Office of Mgmt. & Budget, to Heads of Dep’ts & Agencies 1 (Jun. 22, 2009); Memorandum from Robert Shea, Assoc. Dir., Office of Mgmt. & Budget, to Fed. Agencies 3 (Mar. 6, 2008).

misreported spending in 2009 alone.”¹⁷⁹

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. The use of such algorithms for regulatory enforcement purposes raises a host of novel questions about agency delegations, justifications, and reasoning.¹⁸⁰ Moreover, increased skepticism of the accuracy and objectivity of algorithmic decisionmaking warrants further investigation, particularly as regulators rely on these methods more.¹⁸¹

A related criticism of agency databases is that they often present data without appropriate context.¹⁸² 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.¹⁸³ 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.

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.¹⁸⁴ Such arguments were made repeatedly by industry as the CFPB built its Consumer Complaint Database. Commenters objected that the database would necessarily include only self-selected complaints that were non-random and thus not representative of the consumer population.¹⁸⁵ Industry commenters also

¹⁷⁹ Transparency Through Technology: Evaluating Federal Open-Government Initiatives: Hearing Before the Subcomm. on Tech. of the H. Comm. on Gov’t Reform and Oversight, 112th Cong. 3 (2011) (statement of Ellen Miller, Executive Director, Sunlight Foundation); Shkabatur, *supra* note __, at 103.

¹⁸⁰ See, e.g., Mariano-Florentino Cuellar, *Cyberdelegation and the Administrative State*, (Oct. 1, 2016), at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2754385.

¹⁸¹ 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).

¹⁸² See, e.g., Shkabatur, *supra* note __, 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).

¹⁸³ Shkabatur, *supra* note __, at 104-05 (noting that these are two real-life examples in USASpending.gov).

¹⁸⁴ Shkabatur, *supra* note __, at 117.

¹⁸⁵ See, e.g., 77 Fed. Reg. at 37,561.

objected that the complaints lacked context, and that users might overlook the data's limitations, despite disclaimers by the Bureau.¹⁸⁶

Another problem with agency disclosure, ironically, is the volume of it, and the risk of drowning the public with information. Data.gov currently hosts over 195,000 data sets,¹⁸⁷ some of which include millions of records or data points. But more information is not always better information—although many disclosure regimes assume otherwise.¹⁸⁸ Today, federal agencies post so much information online today that many scholars wonder who is served by these data dumps.¹⁸⁹ 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.”¹⁹⁰ Genuinely useful information is often buried.¹⁹¹ As Paredes notes, sunlight can be the best disinfectant, “[b]ut sunlight can also be blinding.”¹⁹²

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.”¹⁹³ Gradually, because mandated disclosures are so attractive to policymakers, they tend to accumulate over time, which only compounds the problem—“disclosures are added, never removed.”¹⁹⁴ As Ben-Shahar and Schneider observe, consumers “encounter too many disclosures to digest most of them.”¹⁹⁵ 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.”¹⁹⁶

A related shortcoming of disclosure is the complexity of the data. Mandated disclosure regimes have become ubiquitous in federal and state law, ensconced

¹⁸⁶ See, e.g., 77 Fed. Reg. at 37,562.

¹⁸⁷ Data.gov, <http://www.data.gov> (last visited February 14, 2017).

¹⁸⁸ Ben-Shahar & Schneider, *supra* note __, at 650.

¹⁸⁹ Shkabatur, *supra* note __, at 118.

¹⁹⁰ Vladeck, *supra* note __, at 1832.

¹⁹¹ Ben-Shahar & Schneider, *supra* note __, at 737.

¹⁹² Paredes, *supra* note __, at 417.

¹⁹³ 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 Right to Remain Silent”: *Miranda After Twenty Years*, 55 AM. SCHOLAR 367, 368 (1986)); Ben-Shahar & Schneider, *supra* note __, at 678.

¹⁹⁴ Ben-Shahar & Schneider, *supra* note __, at 679.

¹⁹⁵ Ben-Shahar & Schneider, *supra* note __, at 705.

¹⁹⁶ Shkabatur, *supra* note __

in statutes, ordinances, agency regulations, and common law.¹⁹⁷ Sometimes, these sources of law demand “marvelously elaborate disclosures” that are difficult if not impossible for the intended beneficiaries (usually consumers) to understand.¹⁹⁸ Consumers, of course, are not perfectly rational, but exercise “bounded rationality” due to various cognitive biases and distortions.¹⁹⁹ 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....”²⁰⁰ 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.²⁰¹

Given the volume and complexity of most data, intended beneficiaries often do not and cannot use it, particularly in the 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.”²⁰² 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,²⁰³ which often goes unread.²⁰⁴ 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.”²⁰⁵ 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. But successful disclosure systems often require “a distinctive and demanding architecture.”²⁰⁶ Any thoughtful disclosure regime

¹⁹⁷ Ben-Shahar & Schneider, *supra* note __, at 650.

¹⁹⁸ Ben-Shahar & Schneider, *supra* note __, at 650.

¹⁹⁹ FUNG ET AL., *supra* note __, 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 (1955); HERBERT A. SIMON, *MODELS OF BOUNDED RATIONALITY: ECONOMIC ANALYSIS AND PUBLIC POLICY* (1982).

²⁰⁰ FUNG ET AL., *supra* note __, at 34.

²⁰¹ Craswell, *supra* note __, at 361.

²⁰² Ben-Shahar & Schneider, *supra* note __, at 665.

²⁰³ The Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act, Pub. L. No. _____, *codified at* 20 U.S.C. § 1092(f).

²⁰⁴ Ben-Shahar & Schneider, *supra* note __, at 702-04; Sara Lipka, *Do Crime Statistics Keep Students Safe?*, CHRON. HIGHER EDUC., Jan. 30, 2009, at A1;

²⁰⁵ Office of Postsecondary Educ., U.S. Dep’t of Educ., *The Handbook for Campus Crime Reporting* 3 (2005); Ben-Shahar & Schneider, *supra* note __, at 702-03.

²⁰⁶ FUNG ET AL., *supra* note __, at 39.

must determine what information must be disclosed, by whom, to whom, the optimal format, and appropriate quality assurance and enforcement mechanisms.²⁰⁷ Indeed, many assume that disclosure avoids many of the compliance and enforcement costs that attend to traditional command-and-control regulation. But compliance with disclosure regimes must also be monitored and enforced—frequently with both civil and criminal penalties.²⁰⁸ For regimes with low rates of compliance, policymakers have often increased both penalties and enforcement efforts.²⁰⁹

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.²¹⁰ Another recent study found that U.S. physicians in just four common specialties spend \$15.4 billion annually reporting under various quality measurement programs.²¹¹ 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.²¹² Traditional command-and-control regulation long ago lost its luster among policymakers and academics.²¹³ 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.”²¹⁴ One might reasonably wonder whether the time and personnel agencies spend on disclosure would be better spent writing regulations and enforcing them.

²⁰⁷ FUNG ET AL., *supra* note __, at 39-46.

²⁰⁸ FUNG ET AL., *supra* note __, at 45 (examining civil and criminal penalties to enforce the disclosure mandates in campaign finance law and corporate law).

²⁰⁹ FUNG ET AL., *supra* note __, at 45-46.

²¹⁰ SEC, Form 8-K Disclosure of Certain Management Transactions, Securities Act Release No. 33-8090, 77 SEC Dock. 1072 (Apr. 12, 2002).

²¹¹ 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. CMS, National Health Expenditure Data, Historical, at <https://www.cms.gov/research-statistics-data-and-systems/statistics-trends-and-reports/nationalhealthexpenddata/nationalhealthaccountshistorical.html>.

²¹² Ben-Shahar & Schneider, *supra* note __, at 737-42.

²¹³ 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).

²¹⁴ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1348 (citing Ripken, *supra* note __).

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.”²¹⁵ There is also evidence that the party required to disclose information often interprets their compliance with the disclosure requirement as granting them license to act more harshly.²¹⁶ 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 web sites host hundreds of thousands of agency databases (as of February 2017, over 195,000).²¹⁷ As a result, writing about government data is difficult because it has become so very common.²¹⁸ Similarly, a comprehensive survey of agency databases would “pointlessly burden” the audience—risking the same information overload often imposed by policymakers that rely on disclosure.²¹⁹ Nevertheless, as becomes quickly obvious,²²⁰ agency databases are becoming the norm rather than the exception.

Below I survey six of the most salient agency efforts to post information online in searchable formats. Of course, there are many, many more worth discussion—too many to cover in 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 the underlying 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 corruption and the appearance of it.²²¹ Indeed, disclosure has been perhaps the one leg of the campaign finance law tripod to be spared by the

²¹⁵ Ben-Shahar & Schneider, *supra* note __, at 740.

²¹⁶ Ben-Shahar & Schneider, *supra* note __, at 739 (citing such evidence related to conflict-of-interest disclosures).

²¹⁷ Data.gov, at <http://www.data.gov> (“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.gov, Search for a Dataset, http://catalog.data.gov/dataset#sec-organization_type (See “Organization Types” in the left sidebar).

²¹⁸ Ben-Shahar & Schneider, *supra* note __, at 652.

²¹⁹ *Id.* at 652.

²²⁰ Cortez, ACUS Report, Appendix E.

²²¹ *Buckley v. Valeo*, 424 U.S. 1, 66-68 (1976).

Supreme Court. Unlike limits on campaign contributions and expenditures, disclosure requirements have endured repeated First Amendment challenges—from early cases like *Burroughs*²²² to more strident recent cases like *Citizens United* and *McCutcheon*.²²³ Of course, in the seminal case *Buckley v. Valeo*,²²⁴ the Court upheld various disclosure requirements introduced in the Federal Election Campaign Act of 1971 (FECA) and the FECA Amendments of 1974.²²⁵ 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.²²⁶ 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.²²⁷

Today, the FEC maintains several searchable online databases on its web site, which are centralized on the FEC’s Campaign Finance Disclosure Portal.²²⁸ Users can search FEC data based on reports required of federal candidates, parties, committees, donors, and lobbyists, among others.²²⁹ The data are generally searchable by name, date, and location,²³⁰ and are presented in list, map, and chart form.²³¹ 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.²³² The FEC’s Disclosure Data Catalog publishes these datasets in

²²² *Burroughs v. United States*, 290 U.S. 534 (1934) (upholding an indictment under the Federal Corrupt Practices Act).

²²³ *Citizens United v. FEC*, 558 U.S. 310, 319 (2010); *McCutcheon v. FEC*, 134 S.Ct. 1434, 1442, 1459-60 (2014).

²²⁴ 424 U.S. 1 (1976).

²²⁵ 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).

²²⁶ 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).

²²⁷ See, e.g., Lucian A. Bebchuk & Robert J. Jackson, Jr., *Shining Light on Corporate Political Spending*, 101 GEO. L.J. 923 (2013).

²²⁸ FEC, Campaign Finance Disclosure Portal, <http://www.fec.gov/pindex.shtml>.

²²⁹ FEC, Campaign Finance Reports and Data, <http://www.fec.gov/disclosure.shtml>. The FEC’s new beta site is a work in progress. See FEC, Campaign Finance Data (beta site), at <https://beta.fec.gov/data/>.

²³⁰ FEC, Searchable Presentations, <http://www.fec.gov/portal/searchable.shtml>.

²³¹ See, e.g., FEC, 2016 Presidential Campaign Finance, <http://www.fec.gov/disclosure/pnational.do>.

²³² FEC, Disclosure Data Catalog, <http://www.fec.gov/data/DataCatalog.do?cf=searchable>.

downloadable .CSV, .XML, or .XSD formats.²³³

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 two-year summary of the money they raised and spent.²³⁴

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.²³⁵ The system includes, for example, a searchable and sortable list of parties required to pay administrative fines for violating reporting requirements.²³⁶ 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.

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.²³⁷ The Act requires House and Senate officials to use “computerized systems” with “coding” and “cross-indexing” to “maximize public access to materials filed.”²³⁸ Congress also requires these reports to be available over the Internet.²³⁹ 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.²⁴⁰ The 2007 amendments also linked lobbying information with campaign contribution data reported to the FEC,²⁴¹ so that users can better track financial ties between lobbyists and public officials.

²³³ FEC, Disclosure Data Catalog, <http://www.fec.gov/data/DataCatalog.do>.

²³⁴ See, e.g.,

²³⁵ FEC, FEC Enforcement Query System (EQS), at <http://eqs.fec.gov/eqs/searcheqs>.

²³⁶ FEC, Administrative Fines, <http://www.fec.gov/data/AdminFine.do?format=html>.

²³⁷ Pub. L. No. 104-65, 109 Stat. 691, codified as amended at 2 U.S.C. §§ 1601-07.

²³⁸ *Id.* at § 6.

²³⁹ *Id.* at § 6(a)(4), (9).

²⁴⁰ Honest Leadership and Open Government Act of 2007 § 103(b), Pub. L. No. 110-81, 121 Stat. 735, 739. See also Office of the Clerk, U.S. House of Representatives, Lobbying Disclosure, at <http://lobbyingdisclosure.house.gov>; U.S. Senate, Query the Lobbying Disclosure Act Database, at <http://soprweb.senate.gov/index.cfm?event=selectFields&reset=1>.

²⁴¹ Honest Leadership & Open Government Act, *supra* note __, at § __.

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.²⁴² 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.²⁴³ To note just one example, the FEC tracks campaign contributions, but does not track individual contributors well because there is no unique identifier assigned to them.²⁴⁴ Thus, the data “gives the illusion of transparency, but functions instead to obscure the most pertinent financial constituencies in a sea of data.”²⁴⁵ These design failures mean that the intended users—voters, intermediaries, and regulators—are relying on flawed data.²⁴⁶

Scholars have scrutinized at length whether voluminous campaign finance data succeeds in achieving its stated goals, and have identified a long list of preconditions necessary 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.²⁴⁷ The preconditions are substantially similar to the conditions prescribed by Fung, Graham, and Weil (which I discuss in more detail in Part V, *infra*).²⁴⁸ 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.²⁴⁹ Simple fixes, like assigning unique identifiers to contributors, and allowing the FEC to conduct random audits, might improve the quality of the data greatly.²⁵⁰

Finally, as with other data regimes, intermediaries play an important role in translating and synthesizing the government’s campaign finance data. Groups

²⁴² Heerwig & Shaw, *supra* note __. In particular, 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).

²⁴³ Heerwig & Shaw, *supra* note __, at 1479-85.

²⁴⁴ Heerwig & Shaw, *supra* note __, at 1484.

²⁴⁵ Heerwig & Shaw, *supra* note __, at 1484.

²⁴⁶ Heerwig & Shaw, *supra* note __, at 1486-89.

²⁴⁷ Michael J. Malbin & Thomas L. Gais, *The Day After Reform: Sobering Campaign Finance Lessons from the American States* 36 (1998).

²⁴⁸ Heerwig & Shaw, *supra* note __, at 1475 (citing Fung *et al.*, *supra* note __, at 11).

²⁴⁹ For just a small sampling, see, e.g., Heerwig & Shaw, *supra* note __; Hasen, *supra* note __; Richard Briffault, *Updating Disclosure for the New Era of Independent Spending*, 27 J.L. & POL. 683 (2012).

²⁵⁰ Heerwig & Shaw, *supra* note __, at 1494-99.

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.²⁵¹ Indeed, an old 2002 survey found that political journalists relied on the CRP web site more than any other, with the FEC's own site ranking third.²⁵² Another group called MAPlight.org mashes up the publicly-available voting records of members of Congress with campaign finance data, trying to find correlations.²⁵³ Maplight gathers data from not only the FEC, but also from the Center for Responsive Politics, which runs OpenSecrets.org, and FollowTheMoney.org (for California data).²⁵⁴ 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.²⁵⁵

Thus, although there is low confidence in the FEC's campaign finance 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 should enable better, more frequent uses of the information and perhaps even affect disclosers' behavior.

B. *The EPA's Toxic Release Data*

Like campaign finance law, environmental law relies on public disclosure to regulate underlying conduct. Scholars have long recognized this dynamic, which they sometimes refer to as "regulation by revelation"—leveraging the threat of public backlash to change the underlying behavior that leads to pollution.²⁵⁶ As EPA's former General Counsel observed, "Information ... can be a supplement, sometimes even an alternative, to regulation. When broadly available, information can change behavior."²⁵⁷

This logic undergirds several well-known environmental statutes that are predicated on disclosure, such as the Clean Air Act, the Emergency Planning and

²⁵¹ Heerwig & Shaw, *supra* note __ 1487-88.

²⁵² Heerwig & Shaw, *supra* note __, at 1487-88 (citing Albert L. May, *The Virtual Trail: Political Journalism on the Internet*, INST. FOR POLITICS, DEMOCRACY, & THE INTERNET 22 (2002), at http://www.pewtrusts.org/uploadedFiles/wwwpewtrustsorg/Reports/The_practice_of_journalism/pp_online_journalist.pdf).

²⁵³ Maplight.org, <http://www.maplight.org>.

²⁵⁴ Maplight.org, Find Contributions, About or Data (at the bottom of the page), <http://maplight.org/us-congress/contributions>.

²⁵⁵ Heerwig & Shaw, *supra* note __, at 1490.

²⁵⁶ Esty, *supra* note __, 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).

²⁵⁷ Environmental Law Institute, *The Environmental Forum* 36 (July/August 1998).

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.²⁵⁸ The commonality is that these statutes all “place affirmative duties on federal agencies to make information available to the public.”²⁵⁹ In combination, these laws “seem to provide a right of public access to virtually all environmental information in the hands of the federal government.”²⁶⁰

The most well-known EPA 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.”²⁶¹ 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.²⁶² The EPA first reported TRI data in 1989, and first published it online in 1998.²⁶³ 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.²⁶⁴ The data are presented in colorful chart, graph, and map forms.²⁶⁵ For each reporting facility, the EPA maintains a “Facility Profile Report” with more granular data regarding the amount of chemicals managed, released, or transferred. Still more data is available by link to each company’s full reports in the EPA’s Envirofacts database.²⁶⁶

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.²⁶⁷ The initial success of the TRI program

²⁵⁸ David C. Vladeck, Information Access—Surveying the Current Legal Landscape of Federal Right-to-Know Laws, 86 TEX. L. REV. 1787, 1788 (2008) (citations omitted).

²⁵⁹ *Id.* at 1788.

²⁶⁰ *Id.* at 1788.

²⁶¹ Pub. L. No. 99-499 § 313, codified as amended at 42 U.S.C. §§ 11023(h), (j).

²⁶² 30 C.F.R. § 372.65; *see also* 42 U.S.C. § 11023.

²⁶³ EPA, TRI Program Milestones, at http://www.epa.gov/sites/production/files/documents/tritimeline_0.pdf.

²⁶⁴ EPA, Toxics Release Inventory (TRI) Program, <http://www.epa.gov/toxics-release-inventory-tri-program>.

²⁶⁵ *See, e.g.*, EPA, 2014 TRI Factsheet: City – Dallas, TX (2014 Dataset), at https://iaspub.epa.gov/triexplorer/tri_factsheet.factsheet?&pstate=TX&pcity=Dallas&pyear=2014&pDataSet=TRIQ1.

²⁶⁶ EPA, Envirofacts, at <https://www3.epa.gov/enviro/index.html>.

²⁶⁷ Mark A. Cohen & W. Kip Viscusi, *The Role of Information Disclosure in Climate Mitigation Policy*, 3 (4) CLIMATE CHANGE ECO. 1250020-1, 1250020-2 (2012).

even surprised the EPA and environmental groups, who had toiled for years to regulate toxic pollution.²⁶⁸ Ten years after initiating the TRI program, the amount of pollution released had dropped by half.²⁶⁹ 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.²⁷⁰ As one of the earliest programs of its kind, the TRI has been applauded by many as the best example of regulation via disclosure.²⁷¹

Perhaps inspired by its own success, the EPA now publishes hundreds of datasets online. In fact, the EPA web site publishes so many datasets that it includes several landing pages that help users search for and navigate the data available.²⁷² Some of the more well-known datasets after TRI include the EPA's Enforcement and Compliance History Online (ECHO) web site, which allows users to search for recent and historical enforcement actions, including the last date of inspection.²⁷³ The EPA currently lists 96 datasets on its web site,²⁷⁴ with 1,738 listed on Data.gov.²⁷⁵ The agency even hosts an online discussion forum for data developers.²⁷⁶

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.²⁷⁷ They found, in particular, that toxic release data is not embedded in potential users' decisionmaking, as “[m]ost home buyers, renters, job seekers, consumers, and investors do not consider toxic pollution” when making decisions that might be affected by it.²⁷⁸ The TRI data did succeed in better informing policymakers in Congress and the EPA.²⁷⁹ And many manufacturers quickly embedded the new data into their

²⁶⁸ FUNG ET AL., *supra* note __, at 29.

²⁶⁹ FUNG ET AL., *supra* note __, at 29.

²⁷⁰ FUNG ET AL., *supra* note __, at 29.

²⁷¹ See, e.g., Esty, *supra* note __; Pederson, *supra* note __.

²⁷² See, e.g., EPA, Environmental Dataset Gateway (EDG), at <https://edg.epa.gov/metadata/catalog/main/home.page>; EPA, Envirofacts, at <https://www3.epa.gov/enviro/>; EPA, Data Finder, at <https://iaspub.epa.gov/enviro/datafinder.agency>.

²⁷³ EPA, Enforcement and Compliance History Online (ECHO), at <https://echo.epa.gov>.

²⁷⁴ EPA, Data – Developer Central, at <https://developer.epa.gov/category/data/>.

²⁷⁵ Data.gov, U.S. Environmental Protection Agency, at http://catalog.data.gov/dataset?organization=epa-gov&_organization_limit=0 (last accessed on Mar. 11, 2016).

²⁷⁶ EPA, Data and Developer Forum, at <http://blog.epa.gov/data/>. The forum does not appear to be particularly active.

²⁷⁷ FUNG ET AL., *supra* note __, at 76, 79.

²⁷⁸ FUNG ET AL., *supra* note __, at 85.

²⁷⁹ FUNG ET AL., *supra* note __, at 86.

decisionmaking, recognizing the reputational and regulatory consequences.²⁸⁰ 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.²⁸¹ Scholars have long worried that EPA datasets are "patchy" and "unreliable."²⁸²

Moreover, in the last decade, despite its success, the EPA has "drastically scaled back the information made public" under the TRI program.²⁸³ Prior to 2006, the EPA required facilities to report information regarding any chemical release over 500 pounds.²⁸⁴ 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.²⁸⁵ The GAO criticized the EPA's rule as reducing the "quantity and detail of information" released to communities.²⁸⁶ Under a Trump Administration, one could 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."²⁸⁷ 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.²⁸⁸

Of course, even data optimists believe that there are significant challenges in producing data that is reliable and usable.²⁸⁹ But, as in other fields, environmental scholars see great promise in third party watchdogs and data intermediaries translating voluminous data into usable heuristics for consumers, creating a race to the top among regulated firms.²⁹⁰

²⁸⁰ FUNG ET AL., *supra* note __, at 86.

²⁸¹ FUNG ET AL., *supra* note __, at 86 (citing studies).

²⁸² Esty, *supra* note __, at 156.

²⁸³ Vladeck, *supra* note __, at 1791.

²⁸⁴ EPA, 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).

²⁸⁵ *Id.* at 76,937.

²⁸⁶ U.S. Gov't Accountability Office, GAO-08-115T, Environmental Right-to-Know: EPA's Recent Rule Could Reduce Availability of Toxic Chemical Information Used to Assess Environmental Justice 16 (2007), at <http://www.gao.gov/new.items/d08115t.pdf>; Vladeck, *supra* note __, at 1791 n.33.

²⁸⁷ Esty, *supra* note __, at 119.

²⁸⁸ Esty, *supra* note __, at 119.

²⁸⁹ Esty, *supra* note __, at 171-74.

²⁹⁰ Esty, *supra* note __, at 208-09.

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.”²⁹¹ In short, the CFPB “is a new agency operating under a new statute and is on the frontier of the open data trend.”²⁹² 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.²⁹³

The CFPB web site allows consumers to file complaints for 11 categories of financial products, including mortgages, student loans, and credit cards.²⁹⁴ 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.²⁹⁵ The published data is searchable and sortable, 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 consumer.²⁹⁶ Companies select 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.”²⁹⁷ Companies have a total of 60 days to respond, and late responses are tagged by the CFPB as “Past due” or “No response” if the delay exceeds 30 days.²⁹⁸ Each complaint and response is published, but only if they meet numerous publication criteria.²⁹⁹

²⁹¹ Katherine Porter, *The Complaint Conundrum: Thoughts on the CFPB's Complaint Mechanism*, 7 BROOK. J. CORP. FIN. & COM. L. 57, 86 (2012).

²⁹² Cortez, ACUS Report, *supra* note __, 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).

²⁹³ CFPB, Consumer Complaint Database, at <http://www.consumerfinance.gov/complaintdatabase/>.

²⁹⁴ *Id.* The database initially posted complaints only for credit cards. See CFPB, 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. CFPB, Disclosure of Consumer Complaint Data, 77 Fed. Reg. 37,616 (Jun. 22, 2012).

²⁹⁵ 77 Fed. Reg. at 37,616-17.

²⁹⁶ Cortez, ACUS Report, *supra* note __, at 62-63.

²⁹⁷ CFPB, Company Portal Manual (Version 2.14) (May 2015), <http://www.cfjblaw.com/files/Uploads/Documents/Resources/cfpb-company-portal-manual-may-2015.pdf>.

²⁹⁸ *Id.* at 24.

²⁹⁹ The Bureau will not publish complaints if they are missing critical information, have been

In 2014, the Bureau expanded the database (and reversed its previous position) by posting narrative comments by consumers.³⁰⁰ It originally declined to do so, citing privacy and the risk of disclosing consumers' personal information.³⁰¹ 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.³⁰² As a measure of symmetry, the Bureau proposed to allow companies to post their own narrative responses,³⁰³ but companies preferred to respond with a pre-set list of "structured" responses, such as "Company acted appropriately," "Factual dispute," "Misunderstanding," and "Opportunity for improvement."³⁰⁴ However, these responses are optional; companies need not select one for publication.³⁰⁵ 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 relatively sanitized data alone.³⁰⁶

The intended users of the Complaint Database are consumers, researchers, the Bureau, other regulators, and even the subjects of the complaints themselves—companies.³⁰⁷ The Bureau and consumers groups emphasize that the primary intended beneficiaries are consumers, and that publishing complaint data is a "public service" that can "empower" consumers and help them avoid "bad actors" in these markets.³⁰⁸ The Bureau itself emphasizes that "disclosure is one of the best tools government agencies can use."³⁰⁹ 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."³¹⁰ Bureau staff also hope that intermediaries develop mobile apps and

referred to other regulators, are duplicative, would reveal trade secrets, are fraudulent, or identify the wrong company. *Id.* at 26; CFPB, Final Policy Statement: Disclosure of Consumer Complaint Data, 78 Fed. Reg. 21,218, 21,225 (Apr. 10, 2013).

³⁰⁰ CFPB, Disclosure of Consumer Complaint Narrative Data, Notice of Proposed Policy Statement with Request for Public Comment, 79 Fed. Reg. 42,765 (Jul. 23, 2014).

³⁰¹ 77 Fed. Reg. at 37,568.

³⁰² CFPB, Final Policy Statement: Disclosure of Consumer Complaint Narrative Data, 80 Fed. Reg. 15,572 (Mar. 24, 2015).

³⁰³ 79 Fed. Reg. at 42,768.

³⁰⁴ Company Portal Manual, *supra* note __, at 29.

³⁰⁵ 80 Fed. Reg. at 15,583.

³⁰⁶ Cortez, ACUS Report, *supra* note __, at 65.

³⁰⁷ Cortez, ACUS Report, *supra* note __, at 65; 76 Fed. Reg. at 76,630-31.

³⁰⁸ 78 Fed. Reg. at 21,220.

³⁰⁹ 78 Fed. Reg. at 21,220.

³¹⁰ Richard Cordray, Director, CFPB, Remarks at the Consumer Response Field Hearing (Mar. 28, 2013), <http://www.consumerfinance.gov/speeches/prepared-remarks-of-director-richard->

other information products based on complaint data.³¹¹

This latter aspiration that intermediaries will use the government’s data is being realized, at least modestly. Academics are publishing empirical analyses of the CFPB’s data.³¹² Public interest research groups (PIRGs) are producing reports of certain financial product categories, like credit cards and debt collection, based on Bureau data.³¹³ In 2015, the rankings-crazed *U.S. News & World Report* ranked credit cards by relying, in part, on data from the Consumer Complaint Database.³¹⁴

A parallel with other agency data efforts is that the CFPB hopes that the act of publication itself will encourage companies to improve their underlying behavior.³¹⁵ Bureau staff report that complaints have inspired some companies, for example, to address potential problems of their own, such as long customer service phone trees.³¹⁶ Indeed, one of the Bureau’s original aspirations was that published complaints would encourage companies to compete upward in a race to the top, based on how they handled customer service and customer complaints.³¹⁷ Management consulting firms now advise companies to “turn what they hear from the CFPB’s consumer complaint database into a business advantage.”³¹⁸ Bureau staff also report that some companies have tied executive compensation to how well the company has responded to published complaints.³¹⁹ In short, the Bureau sees its role somewhat modestly as a publisher—providing a window into a dialogue between companies and their customers.³²⁰ Once complaints are published, “The marketplace of ideas then does the rest.”³²¹

Of course, the Consumer Complaint Database has not been without controversy,

[cordray-at-the-consumer-response-field-hearing/](#).

³¹¹ Cortez, ACUS Report, *supra* note __, at 66.

³¹² See, e.g., Ian Ayres, Jeff Lingwall, & Sonia Steinway, *Skeletons in the Database: An Early Analysis of the CFPB’s Consumer Complaints*, 19 FORDHAM J. CORP. & FIN. L. 343 (2014).

³¹³ See, e.g., U.S. PIRG, REPORTS: THE CFPB GETS RESULTS FOR CONSUMERS, at <http://www.uspirg.org/page/usf/reports-cfpb-gets-results-consumers>.

³¹⁴ U.S. News & World Report, *How U.S. News Ranks the Best Credit Cards*, <http://money.usnews.com/money/personal-finance/articles/2015/08/13/how-us-news-ranks-the-best-credit-cards?page=3> (Aug. 13, 2015).

³¹⁵ Cortez, ACUS Report, *supra* note __, at 66.

³¹⁶ Cortez, ACUS Report, *supra* note __, at 66.

³¹⁷ 76 Fed. Reg. at 76,630 n.9.

³¹⁸ Deloitte, Analysis: CFPB’s Consumer Complaint Database: Deloitte’s Analysis Reveals Valuable Insights, <http://www2.deloitte.com/us/en/pages/financial-services/articles/consumer-financial-protection-bureau-cfpb-consumer-complaint-database.html>.

³¹⁹ Cortez, ACUS Report, *supra* note __, at 66.

³²⁰ Cortez, ACUS Report, *supra* note __, at 66-67.

³²¹ 76 Fed. Reg. at 76,631.

particularly to the firms identified therein. 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 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 seem to be endorsed by the Bureau; (iv) the data is 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.³²² 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.³²³

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.³²⁴ 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 the public complaint database."³²⁵

Thus, as a new database being run by a new agency, the database continues to be refined and new uses (and objections) continue to emerge. Although the Consumer Complaint Database seems well-designed to defuse potential criticisms, it is still susceptible to being undermined by an unsympathetic Trump Administration and a Republican-led Congress.

D. The CPSC's Product Safety Data

The CPSC's database, SaferProducts.gov, enjoys more clear statutory authority and design 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."³²⁶ But in 2008, Congress passed the

³²² Cortez, ACUS Report, *supra* note __, at 67-70.

³²³ Cortez, ACUS Report, *supra* note __, at 67-71.

³²⁴ Cortez, ACUS Report, *supra* note __, at 75.

³²⁵ Federal Reserve, OIG, Work Plan (Current as of June 5, 2015), <http://oig.federalreserve.gov/reports/work-plan-full.htm#CFPBOngoing> ("Audit of the CFPB's Public Consumer Complaint Database").

³²⁶ 15 U.S.C. § 2051(b).

Consumer Product Safety Improvement Act, requiring the Commission to create a searchable online database of product safety incidents.³²⁷ 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.³²⁸ Each report must describe the product or substance at issue, identify the manufacturer or labeler, and describe the harm reported.³²⁹ 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.”³³⁰ The CPSC launched SaferProducts.gov in March 2011.³³¹

Today, SaferProducts.gov includes a searchable online database of thousands “reports of harm” related to identified products.³³² (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.³³³) The database is keyword-searchable, with advanced search options that include product name, company or brand name, and the product model.³³⁴ 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.³³⁵

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.”³³⁶ As such, SaferProducts.gov includes a disclaimer that tracks this language almost verbatim.³³⁷

³²⁷ Consumer Product Safety Improvement Act of 2008 (CPSIA) § 212, Pub. L. No. 110-314, 122 Stat. 3016 (codified at 15 U.S.C. 2055a).

³²⁸ 15 U.S.C. § 2055a(b)(1)(A).

³²⁹ 15 U.S.C. § 2055a(b)(2)(B). CPSC regulations provide more detail. 16 C.F.R. § 1102.10(d).

³³⁰ 154 Cong. Rec. H7577-01 (2008) (statement by Rep. Eshoo); E1645-01 (2008) (statement by Rep. Delauro); H7577-01 (2008) (statement by Rep. Hollen).

³³¹ CPSC published the final rule in 2010. CPSC, Publicly Available Consumer Product Safety Information Database, 75 Fed. Reg. 76832-01 (Dec. 9, 2010) (codified at 16 C.F.R. pt. 1102).

³³² 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* Coalition for Sensible Safeguards, The Consumer Product Incident Database (Aug. 5, 2013), at <http://www.sensible safeguards.org/factsheets/the-consumer-product-incident-database/>.

³³³ The site includes “portals” for both consumers and businesses. *See, e.g.*, CPSC, File a Report, at <https://www.saferproducts.gov/CPSRMSPublic/Incidents/ReportIncident.aspx>; CPSC, Business Portal, at <https://www.saferproducts.gov/CPSRMSPublic/Industry/Home.aspx>.

³³⁴ CPSC, SaferProducts.gov, Advanced Search, at <http://www.saferproducts.gov/Search/AdvancedSearch.aspx>.

³³⁵ *Id.*

³³⁶ 15 U.S.C. § 2055a(b)(5).

³³⁷ CPSC, SaferProducts.gov, at <http://www.saferproducts.gov/Default.aspx> (“CPSC does not guarantee the accuracy, completeness, or adequacy of the contents of the Publicly Available

An innovation in SaferProducts.gov that might be emulated by other agency databases³³⁸ 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 online.³³⁹ Likewise, the CPSC must consider objections that any information in a report is “materially inaccurate,”³⁴⁰ 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.”³⁴¹ 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.³⁴²

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. 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.’”³⁴³ The court sustained the company’s challenge, somewhat remarkably,³⁴⁴ 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.³⁴⁵ The litigation could provide a roadmap for other litigants similarly aggrieved.

Thus, 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 clear how often consumers use the data to make purchasing decisions (perhaps the data is more useful to distributors and others

Consumer Product Safety Information Database on SaferProducts.gov, particularly with respect to information submitted by people outside of CPSC.”).

³³⁸ Cortez, ACUS Report, *supra* note __, at 20-21.

³³⁹ 15 U.S.C. § 2055a(c)(2).

³⁴⁰ 15 U.S.C. § 2055a(c)(4).

³⁴¹ 16 C.F.R. § 1102.26(a)(1).

³⁴² Omnibus Budget Reconciliation Act of 1981, Pub. L. No. 97-35, 95 Stat. 703 (1981); Cortez, ACUS, at 11; James T. O’Reilly, *Libels on Government Websites: Exploring Remedies for Federal Internet Defamation*, 55 ADMIN. L. REV. 507, 542-43 (2003).

³⁴³ *Company Doe v. Tenenbaum*, 900 F. Supp. 2d 572, 597 (D. Md. 2012).

³⁴⁴ Cortez, ACUS Report, at Appendix __ (Table of Federal Cases 197_ - 2014) (surveying similar cases and finding few successful challenges).

³⁴⁵ 900 F. Supp. 2d at 597.

in the supply chain), nor 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.³⁴⁶ The “upbeat” name of the database might have contributed to such misperceptions.³⁴⁷ 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.³⁴⁸

Perhaps third-party informational products could leverage the data more effectively, for example by creating mobile phone apps that could 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 web site allows users to download raw data files, with voluminous links to the agency’s open data plans.³⁴⁹ 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 thousands of non-governmental 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.³⁵⁰

To facilitate such decisions, CMS operates five searchable databases that

³⁴⁶ GAO, Report to Congressional Committees, CPSC: Awareness, Use, and Usefulness of SaferProducts.gov *ii*. (Mar. 2013), at <http://www.gao.gov/assets/660/652916.pdf>.

³⁴⁷ *Id.* at 27; Consumer Reports, Upbeat Name of SaferProducts.gov Confuses Users, at <http://www.consumerreports.org/cro/news/2013/03/upbeat-name-of-saferproducts-gov-confuses-users/index.htm> (discussing the 2013 GAO report).

³⁴⁸ *Id.* at 12-14.

³⁴⁹ CPSC, The CPSC Open Government Plan, at <http://www.cpsc.gov/open>.

³⁵⁰ 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); Mark A. Hall & Carl E. Schneider, Patients as Consumers: Courts, Contracts, and the New Medical Marketplace, 106 MICH. L. REV. 643 (2008).

compare Medicare providers—Hospital Compare,³⁵¹ Physician Compare,³⁵² Nursing Home Compare,³⁵³ Home Health Compare,³⁵⁴ and Dialysis Facility Compare.³⁵⁵ 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 25-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. Hospitals are sortable by distance, by whether they offer emergency services, and by “hospital type.” Home Health facilities are listed by the type of care offered, such as physical therapy and occupational therapy. Physician searches are more complicated (and are probably the least useful of the five), requiring the user to also search for a physician’s name, specialty, or medical condition to help narrow the results.³⁵⁶

The five databases offer quite distinct quality data. Hospital Compare includes data on over 4,000 hospitals nationwide,³⁵⁷ 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.³⁵⁸ 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.³⁵⁹ 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.*,

³⁵¹ CMS, Hospital Compare, at <https://www.medicare.gov/hospitalcompare/search.html>.

³⁵² CMS, Physician Compare, at <https://www.medicare.gov/physiciancompare/search.html>.

³⁵³ CMS, Nursing Home Compare, at <https://www.medicare.gov/nursinghomecompare/search.html>.

³⁵⁴ CMS, Home Health Compare, at <https://www.medicare.gov/homehealthcompare/search.html>.

³⁵⁵ CMS, Dialysis Facility Compare, at <https://www.medicare.gov/dialysisfacilitycompare/#search>.

³⁵⁶ CMS, Physician Compare, About the Data, at <https://www.medicare.gov/physiciancompare/staticpages/data/aboutthedata.html>.

³⁵⁷ CMS, What is Hospital Compare?, at <http://www.medicare.gov/hospitalcompare/About/What-Is-HOS.html>.

³⁵⁸ 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.

³⁵⁹ CMS, Hospital Compare, Footnotes, at <https://www.medicare.gov/hospitalcompare/Data/Footnotes.html>.

“No Different than the National Average Payment” is frequently displayed). Also, “Value of Care” metrics also 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 a 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 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.”³⁶⁰ 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.³⁶¹ 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.³⁶² Physicians who report to CMS are then given performance scores. But the system is voluntary, and only some Physician Compare profiles include their performance scores.³⁶³ Although there are over 200 reportable quality measures, few are listed on Physician Compare.³⁶⁴

³⁶⁰ CMS, 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).

³⁶¹ 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).

³⁶² For a list of examples, see CMS, Physician Compare, Physician Quality Reporting System (PQRS), at <https://www.medicare.gov/physiciancompare/staticpages/data/pqrs.html>.

³⁶³ Id.

³⁶⁴ 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.³⁶⁵

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.³⁶⁶ 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 other many others.³⁶⁷ Today's Compare databases are an amalgam of data from a variety of sources, gradually added like ornaments to a Christmas tree.³⁶⁸

Perhaps the richest potential source of data is Medicare claims data. CMS processes over a billion Medicare claims each year,³⁶⁹ 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.³⁷⁰ The growth of “big data” in health care, in fact, roughly parallels and depends on the growth and accessibility of Medicare claims data.³⁷¹

Although Medicare has released various data to the public for years, CMS was

³⁶⁵ Madison, *supra* note __, at 1625.

³⁶⁶ Madison, *supra* note __, at 1626; CMS, Hospital Compare, at <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html>. 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 tit. IV, subtit. C, codified at

³⁶⁷ Madison, *supra* note __, at 1626 (citing CMS, Hospital Compare, at <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html>).

³⁶⁸ As with other agencies, the raw data files at a separate site. See CMS, Data.Medicare.gov, at <http://data.medicare.gov>.

³⁶⁹ U.S. Dep't of Health & Human Servs., Medicare Fee-for-Service, Payment Accuracy, <http://www.paymentaccuracy.gov/programs/medicare-fee-service>; 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 (2014).

³⁷⁰ Madison, *supra* note __, at 1609-10.

³⁷¹ Madison, *supra* note __, at 1610.

long restricted from releasing physician claims data by court order.³⁷² In 2014, a year after the order was lifted, CMS released claims data for over 880,000 providers.³⁷³ The trend has been followed by states—16 of which have created all-payer claims databases that compile claims data from almost all payers in the state, including public and private insurers.³⁷⁴ However, in 2016 the Supreme Court threw a wrench in these efforts when it interpreted ERISA as preempting state laws mandating reporting from self-insured employer plans.³⁷⁵ As is common for many data regimes, agencies often take two steps forward, then one step back.

Of course, claims data has traditionally been 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,³⁷⁶ and today over 1,300 hospitals report data about infection rates, mortality rates, and other quality indicators in order to boost their Medicare reimbursements,³⁷⁷ 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.³⁷⁸ Yet, realizing that not everyone is covered by Medicare, 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.³⁷⁹ Thus, various data sources are being combined in novel ways.

³⁷² 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).

³⁷³ Madison, *supra* note __, 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>.

³⁷⁴ John D. Freedman *et al.*, All-Payer Claims Databases – Uses and Expanded Prospects after *Gobeille*, 375 (23) N. ENG. J. MED., 2016 WL 37580714 (Dec. 8, 2016).

³⁷⁵ *Gobeille v. Liberty Mutual Ins. Co.*, 577 U.S. __ (2016) (invalidating a Vermont law requiring self-insured employers to submit claims data on grounds of ERISA preemption).

³⁷⁶ 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)).

³⁷⁷ CMS, Medicare Program, 78 Fed. Reg. 50,496, 50,678-80 (Aug. 19, 2013)

³⁷⁸ Madison, *supra* note __, at 1613; CMS, Physician Quality Reporting System (PQRS) Overview 1, 2-3, http://www.cms.gov/Medicare/Quality-Initiatives-Patient-Assessment-Instruments/PQRS/Downloads/PQRS_OverviewFactSheet_2013_08_06.pdf.

³⁷⁹ Madison, *supra* note __, at 1614-15; Patient Protection and Affordable Care Act § 10332, *codified at* 42 U.S.C. § 1395kk(e); CMS, 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; CMS, Expanding Uses of Medicare Data by Qualified Entities, 81 Fed. Reg. 44,456 (Jul. 7, 2016), *codified at* 42 C.F.R. part 401.

Nevertheless, scholars have long questioned the utility of performance data and report cards in the health industry,³⁸⁰ and disclosure more generally has long been a point of interest to health law scholars.³⁸¹ Patients report that they seldom rely on publicly available data or even more comprehensible report cards of physicians, hospitals, or procedures.³⁸² 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.³⁸³ Patients frequently are not aware of the information, do not understand it, or do not use it.³⁸⁴ Indeed, of 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 misinterpreted, and the risk of strategic behavior by providers.³⁸⁵

As a result, CMS has tried to “shrink” the voluminous and varied data by translating them into star ratings, using a five-star scale.³⁸⁶ Five stars represents facilities that are “much above average,” four stars are “above average,” three stars are “average,” and so on. 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.³⁸⁷ For example, a study of cardiac surgery report cards in Pennsylvania found that cardiac surgeons responded to the new disclosure requirement by becoming

³⁸⁰ There is a 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).

³⁸¹ *See, e.g.*, William M. Sage, Regulating Through Information: Disclosure Laws and American Health Care, 99 COLUM. L. REV. 1701.

³⁸² *Id.*

³⁸³ Susannah Fox & Maeve Duggan, Pew Research Ctr., Health Online 2013, 20-23 (Jan. 2013), at http://www.pewinternet.org/files/old-media//Files/Reports/PIP_HealthOnline.pdf.

³⁸⁴ Ben-Shahar & Schneider, *supra* note __, at 672, 711-12 (citing studies).

³⁸⁵ FUNG ET AL., *supra* note __, at 76-77 (analyzing disclosure requirements in Pennsylvania and New York).

³⁸⁶ *See, e.g.*, CMS, Nursing Home Compare, <http://www.medicare.gov/nursinghomecompare/search.html>; Madison, *supra* note __, at 1627.

³⁸⁷ *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, *Regulating Through Information*, *supra* note __, at 1793.

more reluctant to operate on sicker patients.³⁸⁸ 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.³⁸⁹

Moreover, despite the prevalence of doctor and hospital ratings and report cards, it is not clear 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.”³⁹⁰ One study found that less than 1% of patients knew how their hospital or surgeon was rated under mandated ratings systems.³⁹¹ 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 HHS and CMS for mistakenly calculating its star rating on Nursing Home Compare, giving the facility two stars out of five rather than four.³⁹² The mistaken star rating was published on the CMS web site, which did not correct it for almost two years.³⁹³ 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 5th and 14th amendments.³⁹⁴ 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.”³⁹⁵ 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.³⁹⁶ 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.³⁹⁷ 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, or a change in licensing

³⁸⁸ Eric C. Schneider & Arnold M. Epstein, Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery, 279 (20) JAMA 1638 (1998).

³⁸⁹ FUNG ET AL., *supra* note __, at 89 (citing studies)

³⁹⁰ 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), at <http://www.hschange.org/CONTENT/1028/1028.pdf>; Epstein, *Rolling Down the Runway*, *supra* note __, at 1694.

³⁹¹ Epstein, *Rolling Down the Runway*, *supra* note __, at 1694.

³⁹² Bryn Mawr Care v. Sebelius, 898 F. Supp. 2d 1009, 1011 (N.D. Ill. 2012).

³⁹³ *Id.*

³⁹⁴ *Id.* at 1011-12.

³⁹⁵ *Id.* at 1012.

³⁹⁶ *Id.* at 1013.

³⁹⁷ *Id.* at 1014.

status or reimbursement status, or some other tangible harm.³⁹⁸ 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.³⁹⁹

Still, quality ratings and other disclosure-based regulation remains all the rage in health policy,⁴⁰⁰ with contemporary proposals littered with patient surveys, outcomes data, star ratings or rankings, and of course federal databases mentioned above that combine many of these data points.⁴⁰¹ Again, scholars have long been aware of the benefits, burdens, and limitations of using disclosure as a regulatory tool in health care.⁴⁰² Yet, it is possible, if not probable, that efforts to replace the Affordable Care Act⁴⁰³ 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,⁴⁰⁴ Warning Letters,⁴⁰⁵ recalls,⁴⁰⁶ and enforcement reports.⁴⁰⁷ The FDA also maintains a database of good news—agency product approvals.⁴⁰⁸ But perhaps the most 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

³⁹⁸ *Id.* at 1014-15.

³⁹⁹ *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 Nathan Cortez, ACUS Report, *supra* note __, Appendix C: Table of Federal Cases (1974-2014), available at <https://www.acus.gov/sites/default/files/documents/appendix-c.pdf>.

⁴⁰⁰ 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 35(4) HEALTH AFF. (Apr. 2016).

⁴⁰¹ See, e.g., FUNG ET AL., *supra* note __, at 160-62 (listing various disclosure systems in health care).

⁴⁰² See, e.g., Sage, Regulating Through Information, *supra* note __.

⁴⁰³ Pub. L. No. 111-148 (2010); Health Care Education and Reconciliation Act of 2010, Pub. L. No. 111-152 (2010).

⁴⁰⁴ FDA, Inspections Database, at <http://www.fda.gov/ICECI/Inspections/ucm222557.htm>.

⁴⁰⁵ FDA, Inspections, Compliance, Enforcement, and Criminal Investigations: Warning Letters, at <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/>.

⁴⁰⁶ FDA, Recalls, Market Withdrawals, & Safety Alerts, at <http://www.fda.gov/Safety/Recalls/default.htm>.

⁴⁰⁷ FDA, Enforcement Report, at http://www.accessdata.fda.gov/scripts/ires/index.cfm#tabNav_advancedSearch.

⁴⁰⁸ FDA, Drugs@FDA: FDA Approved Drug Products, at <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm>.

medication errors and adverse drug events reported to the agency.⁴⁰⁹ Adverse event reporting dates back at least 35 years, and perhaps even longer, when the agency received reports by paper.⁴¹⁰ Manufacturer reports are required by regulation,⁴¹¹ but reports by health care professionals and consumers is only voluntary.⁴¹²

Evolving from previous iterations,⁴¹³ 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 web site includes aggregate statistics, as well as links to raw data files that include individual case reports.⁴¹⁴ But the raw data files are published only in quarterly increments,⁴¹⁵ 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.")⁴¹⁶ FDA also instructs potential users to request individual case safety reports by submitting a FOIA request.⁴¹⁷ Individual reports are accessible, however, by searching the FDA's MedWatch web site, which aggregates "clinically important safety information" for "human medical products," including drugs, devices, and biologics.⁴¹⁸

⁴⁰⁹ FDA, FDA Adverse Drug Event Reporting System (FAERS),

⁴¹⁰ FDA, Postmarketing Safety Reports for Human Drug and Biological Products; Electronic Submission Requirements, 79 Fed. Reg. 33,072, 33,073 (Jun. 10, 2014). Indeed, one source notes that the FDA has maintained an adverse event database in some form since 1967. Cortez, ACUS Report, *supra* note __, at 44 (citing David Gortler, Adverse Event Databases (AERS Database), at <http://faculty.georgetown.edu/dg298/adverse-event-database.html>).

⁴¹¹ 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).

⁴¹² FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS), at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

⁴¹³ 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; 63 Fed. Reg. 65,000, 65,030 (Nov. 24, 1998).

⁴¹⁴ FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS), at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

⁴¹⁵ FDA, FDA Adverse Event Report System (FAERS), Latest Quarterly Data Files, at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

⁴¹⁶ FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS), at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

⁴¹⁷ *Id.*

⁴¹⁸ FDA, MedWatch: The FDA Safety Information and Adverse Event Reporting Program, at <http://www.fda.gov/Safety/MedWatch/default.htm>. *See also* David A. Kessler *et al.*, *Introducing MEDWatch: A New Approach to Reporting Medication and Device Adverse Effects and Product Problems*, 269

The device counterpart to FAERS is MAUDE (Manufacturer and User Facility Device Experience), a database of device adverse events,⁴¹⁹ which the agency began collecting in 1984.⁴²⁰ 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 10 years.⁴²¹ The MAUDE database allows users to search for medical devices that may have malfunctioned or caused death or serious injury.⁴²² Users can search by a pull-down menu of product problems, by the class of product, or by manufacturer, model, or brand name.⁴²³ 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.⁴²⁴ The site, launched in 2014, includes separate pages for food products, drugs, and devices, with all three including enforcement reports, and the drug and device pages including adverse event databases.⁴²⁵ 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 2004.⁴²⁶ 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.⁴²⁷ The goal was to include at least 100 million patients in

JAMA 2765 (Jun. 2, 1993).

⁴¹⁹ FDA, MAUDE, at <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm>.

⁴²⁰ FDA, Medical Device Reporting, 49 Fed. Reg. 36,326 (Sep. 14, 1984).

⁴²¹ FDA, Mandatory Reporting Requirements: Manufacturers, Importers, and Device User Facilities, at <http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/ReportingAdverseEvents/ucm127891.htm>.

⁴²² Id.

⁴²³ FDA, MAUDE – Manufacturer and User Facility Device Experience, at <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMAUDE/search.CFM>.

⁴²⁴ Cortez, ACUS Report, *supra* note __, at 44; FDA, OpenFDA, at <https://open.fda.gov>.

⁴²⁵ Cortez, ACUS Report, *supra* note __, at 44; FDA, OpenFDA, at <https://open.fda.gov>.

⁴²⁶ FDA, OpenFDA, Drugs – Adverse Events, at <https://open.fda.gov/drug/event/>.

⁴²⁷ Food and Drug Administration Amendments Act of 2007 § 905(a), Pub. L. No. 110-85, 121 Stat. 823, 944 (codified as amended in scattered sections of 21 U.S.C.).

the dataset by 2012.⁴²⁸ Named the FDA Sentinel Initiative, it has been designed to monitor product safety across different data sources, including data from Medicare and the Veterans Health Administration, and large private health insurers.⁴²⁹ A pilot version (called “Mini-Sentinel”) is now available online,⁴³⁰ including a search tool for the Mini-Sentinel Distributed Database.⁴³¹

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% of reports using inaccurate product names.⁴³² More than most agencies, then, the FDA includes prominent disclaimers of the accuracy and reliability of its data.⁴³³ 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.⁴³⁴

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

⁴²⁸ 21 U.S.C. § 355(k)(3)(B).

⁴²⁹ Madison, *supra* note __, 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).

⁴³⁰ FDA, Mini-Sentinel, at <http://mini-sentinel.org>.

⁴³¹ FDA, Mini-Sentinel, Distributed Query Tool and Summary Tables, at http://mini-sentinel.org/data_activities/distributed_query_tool/default.aspx.

⁴³² Kenneth A. Getz, Stella B.A. Stergiopoulos, & Kenneth I. Kaitin, *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).

⁴³³ Cortez, ACUS Report, *supra* note __, at 44.

⁴³⁴ FDA, MAUDE – Manufacturer and User Facility Device Experience, <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm#disclaimer> (last visited July 23, 2015).

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.⁴³⁵

The Open FDA site also confronts users, via pop-up window, with a note of caution that “This API is not for clinical or production use. While we make every effort to ensure that data is accurate, you should assume all results are unvalidated.”⁴³⁶ 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.⁴³⁷ Thus, an alliance between improving FDA and CMS data might help cure some defects in the FDA’s adverse event reporting systems.⁴³⁸ 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,⁴³⁹ 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?⁴⁴⁰ My first prescription is modesty. As Richard Craswell cautions, “people who expect disclosure laws to solve almost every

⁴³⁵ FDA, *FAERS*, *supra* note __.

⁴³⁶ See, e.g., FDA, Open FDA, Drugs – Adverse Events, at <https://open.fda.gov/drug/event/>.

⁴³⁷ See, e.g., Scott Gottlieb, Opening Pandora’s Pillbox: Using Modern Information Tools to Improve Drug Safety, 24 *HEALTH AFF.* 938 (2005).

⁴³⁸ *Id.*

⁴³⁹ For a very broad survey of some of the recent digital technologies, see, e.g., Nathan Cortez, *The Mobile Health Revolution?*, 47 *U.C. DAVIS L. REV.* 1173 (2014).

⁴⁴⁰ These values derive, not coincidentally, from the Information Quality Act, *supra* note __. Note also that the GAO has discussed what it means for data to be “reliable.” See <http://www.gao.gov/assets/80/77213.pdf> at 5.

problem—quickly, easily, and with very little cost—are doomed to have their expectations crushed.”⁴⁴¹ The truth is that regulation by database requires just as many difficult design and implementation choices as any other form of regulation.⁴⁴² The early successes of the CFPB’s Consumer Complaint Database and the CPSC’s SaferProducts.gov site, for example, owe to several canny decisions by Congress and the agencies to ensure the quality of the data posted. It is a mistake, then, to assume that creating a public database is necessarily less difficult and more cost-effective than traditional regulatory enforcement.

Another word of caution is that agency databases vary widely in their purposes, scope, design, sources, and presentation. Although it is neither possible nor wise 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.⁴⁴³ 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).⁴⁴⁴ I thus offer a series of recommendations that address both dimensions, proposing a decidedly modern role for the government as a “data steward.”

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 appropriate 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.⁴⁴⁵ Thus, government agencies can play an important role by ensuring that published data

⁴⁴¹ Craswell, *supra* note __, at 379.

⁴⁴² Craswell, *supra* note __, at 379.

⁴⁴³ ACUS, *supra* note 32.

⁴⁴⁴ 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, Allison K. Hoffman, & William M. Sage, eds.) (2016).

⁴⁴⁵ See, e.g., FUNG ET AL., *supra* note __, at 165 (“Those who contribute information can do so without identifying themselves or their sponsoring organizations, or taking responsibility for what they are saying.”).

is credible.⁴⁴⁶ In fact, because publication by a government agency itself often signals credibility, it is incumbent on agencies to ensure the accuracy of data that purports to be objective.⁴⁴⁷

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.⁴⁴⁸ 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.⁴⁴⁹ Likewise, when Congress authorized SaferProducts.gov, it required the CPSC to consider objections that the information is “materially inaccurate.”⁴⁵⁰

Thus, agencies that purport to publish accurate and objective data should adopt procedures to these ends. As emphasized in Part II, “transparency” is frequently invoked to support disclosure, but it is important to remember that “transparency” allows objects to be “seen without distortion.”⁴⁵¹ Thus, to the extent feasible, government databases should try to achieve genuine transparency rather than “translucency.”⁴⁵² 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.⁴⁵³ But scholars also find that the more successful disclosure regimes tend to increase the accuracy and quality of information they publish over time.⁴⁵⁴ Thus, databases should be dynamic rather than static—a constant work-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.

1. Pre-Publication Procedures

⁴⁴⁶ Heerwig & Shaw, *supra* note __ 1476.

⁴⁴⁷ Cortez, ACUS Report, *supra* note __, at __.

⁴⁴⁸ Madison, *supra* note __, at 1607-08.

⁴⁴⁹ 78 Fed. Reg. 21,218, 21,225-26 (Apr. 10, 2013).

⁴⁵⁰ 15 U.S.C. § 2055(c)(4).

⁴⁵¹ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1343 (referring to the *Oxford English Dictionary*’s definition of “transparency”).

⁴⁵² Schauer, *Transparency in Three Dimensions*, *supra* note __, 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.

⁴⁵³ FUNG ET AL., *supra* note __, at xii.

⁴⁵⁴ FUNG ET AL., *supra* note __, at 109.

Data sets that purport to publish accurate, objective information should include, ideally, both pre- and post-publication procedures that allow parties to request that any information not meeting these standards be corrected and/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.⁴⁵⁵ 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.⁴⁵⁶ The CPSC must consider objections that the information is “materially inaccurate,”⁴⁵⁷ and the Commission publishes these procedures in the C.F.R.⁴⁵⁸ There are clear timelines for parties to object to alleged inaccuracies and for the CPSC to resolve disputes before publication. Likewise, the CFPB authenticates that complaints are coming from actual customers of the company. Bureau procedures allow companies to use an online Company Portal to verify a commercial relationship with the customer, and post the company’s response.⁴⁵⁹ The Bureau also makes clear that each complaint, before being published in the database, must meet several publication criteria.⁴⁶⁰ Of all the databases discussed in Part IV, *supra*, 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.⁴⁶¹

Here, the controversial Information or Data Quality Act might be of help.⁴⁶² The Act required the OMB to publish government-wide guidelines for “ensuring and maximizing the quality, objectivity, utility, and integrity of information ...

⁴⁵⁵ Cortez, ACUS Report, *supra* note __, at 94-95.

⁴⁵⁶ 15 U.S.C. § 2055(c)(2).

⁴⁵⁷ 15 U.S.C. § 2055(c)(4).

⁴⁵⁸ 16 C.F.R. § 1102.26.

⁴⁵⁹ CFPB, *Company Portal Manual* (Version 2.15) (May 2015), <http://www.cfjblaw.com/files/Uploads/Documents/Resources/cfpb-company-portal-manual-may-2015.pdf>. Again, companies can choose among ten pre-set responses. *See* note __, *supra*.

⁴⁶⁰ Cortez, ACUS Report, *supra* note __, at 63-64.

⁴⁶¹ Gellhorn, *supra* note __; Cortez, *supra* note __; Cortez, ACUS Report, *supra* note __ (Appendix C: Table of Federal Cases (1974-2014)).

⁴⁶² The Act has been criticized as an attempt to thwart regulators in the guise of data quality. *See, e.g.,* Levy & Johns, *supra* note __, at 2.

disseminated by the government.”⁴⁶³ The Act applies broadly to “Federal agency dissemination of public information, regardless of the form or format.”⁴⁶⁴ It also directed the OMB to establish procedures that allow “affected persons to seek and obtain correction of information maintained and disseminated by the agency.”⁴⁶⁵ Per the OMB’s guidelines, dozens of federal agencies have published their own such guidelines and post-publication procedures for correcting or retracting information.⁴⁶⁶ 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.”⁴⁶⁷ Second, they exclude “adjudicative processes.”⁴⁶⁸ 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.”⁴⁶⁹ 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.

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 superintending disputes over agency data.⁴⁷⁰

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

⁴⁶³ IQA, *supra* note __. The IQA built on earlier provisions in the Paperwork Reduction Act. Paperwork Reduction Act of 1995, Pub. L. No. 104-13, 109 Stat. 163, 168.

⁴⁶⁴ IQA, *supra* note __; 67 Fed. Reg. 8452, 8460 (Feb. 22, 2002). OMB guidelines define “information” as “any communication or representation of knowledge such as facts or data, in any medium or form,” including “information that an agency disseminates from a web page.” 67 Fed. Reg. at 8460.

⁴⁶⁵ IQA, *supra* note __.

⁴⁶⁶ See Office of Management and Budget, Agency Information Quality Guidelines, https://www.whitehouse.gov/omb/inforeg_agency_info_quality_links/.

⁴⁶⁷ 67 Fed. Reg. at 377.

⁴⁶⁸ *Id.*

⁴⁶⁹ Cortez, ACUS Report, *supra* note __. 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>.

⁴⁷⁰ Cortez, ACUS Report, *supra* note __, at 100-02.

significant information gathering efforts by agencies.⁴⁷¹ 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.⁴⁷² 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 equivalent, which can field complaints about data published by agencies. For example, the CFPB’s Ombudsman has heard complaints about the Consumer Complaint Database,⁴⁷³ and the FDA’s many ombudsmen have fielded complaints under the IQA.⁴⁷⁴ Ombuds can serve important customer service functions with regulated parties. As such, the use of ombuds in federal agencies has gained steam in recent years, as have calls for standards regarding their independence, duties, and information-providing roles.⁴⁷⁵

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.⁴⁷⁶ They also maintain, by design, crucial independence from agency heads, and thus can serve as an independent arbiter.⁴⁷⁷ They can also function as an avenue for fielding industry complaints.⁴⁷⁸ 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.”⁴⁷⁹

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

⁴⁷¹ 44 U.S.C. § 3507(d).

⁴⁷² *Id.* at 3508.

⁴⁷³ CFPB, Ombudsman’s Office, *Annual Report to the Director* (Nov. 17, 2014).

⁴⁷⁴ FDA, FDA’s Office of the Ombudsman: Dispute Resolution and Problem Solving, at 2.

⁴⁷⁵ See, e.g., ACUS, *The Use of Ombuds in Federal Agencies*, at <https://www.acus.gov/recommendation/use-ombuds-federal-agencies> (Dec. 14, 2016).

⁴⁷⁶ Inspector General Act of 1978, Pub. L. No. 95-452, 92 Stat. 1101 (1978).

⁴⁷⁷ 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).

⁴⁷⁸ Cortez, ACUS Report, *supra* note __, at 101.

⁴⁷⁹ Federa. Reserve, OIG, *Work Plan* (Current as of June 5, 2015), <http://oig.federalreserve.gov/reports/work-plan-full.htm#CFPBOngoing> (“Audit of the CFPB’s Public Consumer Complaint Database”).

well to their own data regimes.

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.⁴⁸⁰ Data collection and processing requires not just automation, but human labor. Unfortunately, such labor is dismissed as the task of “data janitors” who receive inadequate compensation.⁴⁸¹ The lack of sufficient infrastructure to ensure data quality can generate “big bad data”⁴⁸²—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 web site is not a simple proposition. Web masters for federal web sites 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.”⁴⁸³ Although each separate requirement may stand on its own logic, together they can limit how agencies present data, and generally favor standardization above experimentation.⁴⁸⁴ 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.”⁴⁸⁵ In this vein, nongovernmental intermediaries have proven useful in rendering government data more accessible and usable,⁴⁸⁶ as I discuss below in section C.

Finally, all this assumes that policymakers have already made the threshold decision to publish the data and accept 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 more optimal behavior from regulated entities? And of existing databases, which low-quality data sets might

⁴⁸⁰ I am indebted to Frank Pasquale for raising many of the issues in this paragraph.

⁴⁸¹ See, e.g., Lilly Irani, Justice for “Data Janitors,” (Jan. 15, 2015), at <http://www.publicbooks.org/nonfiction/justice-for-data-janitors>.

⁴⁸² Sharona Hoffman & Andy Podgurski, Big Bad Data: Law, Public Health, and Biomedical Databases, 41 J. L. MED. & ETHICS 56 (Supp. S1 2013).

⁴⁸³ Robinson *et al.*, *supra* note __, at 162.

⁴⁸⁴ Robinson *et al.*, *supra* note __, at 163-65.

⁴⁸⁵ Robinson *et al.*, *supra* note __, at 165.

⁴⁸⁶ Robinson *et al.*, *supra* note __, at 165.

be worth salvaging? The FEC's and FDA's databases might overwhelm 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.⁴⁸⁷

4. Administrative Law Dimensions

Perhaps most interesting, from an administrative law perspective, is that database publication procedures might be seen 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.⁴⁸⁸ But like traditional agency adjudications, the amount of "procedure" appropriate for database disputes will depend on the unique features of each database, the regulatory and 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 reporting?⁴⁸⁹ Could plaintiffs or prosecutors use the False Reports to the Government Act⁴⁹⁰ or the False Claims Act⁴⁹¹ 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.

⁴⁸⁷ I credit Nick Bagley for seeding the ideas in this paragraph.

⁴⁸⁸ Bill Sage inspired many of the thoughts in this paragraph.

⁴⁸⁹ Discussing this possibility, see Frank Pasquale, *Grand Bargains for Big Data: The Emerging Law of Health Information*, 72 MARYLAND L. REV. 682 (2013).

⁴⁹⁰ 18 U.S.C. § 1001. 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, *Legal and Policy Issues in Measuring and Improving Quality*, *supra* note __, at __.

⁴⁹¹ 31 U.S.C. §§ 3729-33.

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.⁴⁹² 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.⁴⁹³ Thus, the sources of data should be clearly identified, including whether the data are contested and any steps the agency takes (or more importantly does not) to resolve such contests.⁴⁹⁴

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.⁴⁹⁵ 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.⁴⁹⁶ 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.”⁴⁹⁷ 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.⁴⁹⁸ Should such “data” even be published? For better or worse, routinely they are.

⁴⁹² See, e.g., 21 U.S.C. § 343(a) (prohibiting food labeling from being false or misleading).

⁴⁹³ Cortez, ACUS Report, *supra* note __, at 94.

⁴⁹⁴ Cortez, ACUS Report, *supra* note __, at 96.

⁴⁹⁵ Cortez, ACUS Report, *supra* note __, at 96.

⁴⁹⁶ See Part II.F, *supra*.

⁴⁹⁷ Cortez, ACUS Report, *supra* note __, at 69-70; Porter, *The Complaint Conundrum*, *supra* note __, at 78.

⁴⁹⁸ Sandeep Jauhar, *Giving Doctors Grades*, N.Y. TIMES (Jul. 22, 2015), <http://www.nytimes.com/2015/07/22/opinion/giving-doctors-grades.html>. Although death and complication rate data are usually normalized through various methodologies, physicians frequently object that such methodologies are inadequate.

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.”⁴⁹⁹ The FDA’s adverse drug event database also notes that “there is no certainty that the reported event ... was actually due to the product.”⁵⁰⁰ Likewise, the CFPB disclaims that “We don’t verify all the facts alleged in these complaints but we take steps to confirm a commercial relationship between the consumer and the company.”⁵⁰¹ 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.”⁵⁰² Although one court called the CPSC’s language “boilerplate” that “would not interest an ordinary consumer,”⁵⁰³ 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, non-random, and thus non-representative of company performance and consumer experiences.⁵⁰⁴ 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.”⁵⁰⁵

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

⁴⁹⁹ FDA, MAUDE, *supra* note ____.

⁵⁰⁰ FDA, FAERS, *supra* note ____.

⁵⁰¹ CFPB, Consumer Complaint Database, *supra* note __.

⁵⁰² 15 U.S.C. § 2055(b)(5).

⁵⁰³ *Company Doe v. Tenenbaum*, *supra* note __, at 598.

⁵⁰⁴ *See, e.g.*, 77 Fed. Reg. at 37,561.

⁵⁰⁵ *Id.*

accountability.”⁵⁰⁶ 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.”⁵⁰⁷ 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 transparency. 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.⁵⁰⁸ 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 perhaps why.

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 “Most requests for correspondence and other documents are fulfilled by printing them, redacting, then re-scanning into unsearchable images.”⁵⁰⁹ Yet, as far back as 2004, the OMB

⁵⁰⁶ Shkabatur, *supra* note __, at 105-06.

⁵⁰⁷ Adam Samaha, *Government Secrets, Constitutional Law, and Platforms for Judicial Intervention*, 53 UCLA L. REV. 909, 919 (2006); Shkabatur, *supra* note __, at 106.

⁵⁰⁸ *See, e.g.* Vladeck, *supra* note __, at 1828-29.

⁵⁰⁹ The Freedom of Information Act: Ensuring Transparency and Accountability in the Digital Age: Hearing Before the S. Comm. on the Judiciary, 112th Cong. 82 (2011) (statement of Sarah Cohen, Knight Professor of the Practice of Journalism, Duke University).

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.”⁵¹⁰ 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.⁵¹¹ Thus, there is 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 non-programmers and others who are not able to understand or use such data.⁵¹² Raw government datasets might require, ironically, technically sophisticated intermediaries to decipher.⁵¹³ 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 under-performs most other hospitals. But it is much more difficult for users to understand what a 2% 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 “No 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,

⁵¹⁰ Clay Johnson III, Office of Mgmt. & Budget, Executive Office of the President, OMB Memorandum: Policies for Federal Agency Public Websites 4 (2004), at <http://www.whitehouse.gov/omb/memoranda/fy2005/m05-04.pdf>.

⁵¹¹ See, e.g., Robinson *et al.*, *supra* note __, at 167 (arguing that original data should be posted in documents in XML formats with unique and permanent addresses); Open Government Working Group, Open Government Data Principles, <http://wiki.opengovdata.org/index.php/OpenDataPrinciples> (recommending that data be complete, primary, timely accessible, machine-readable, non-discriminatory, non-proprietary, and license-free).

⁵¹² See, e.g., Shkabatur, *supra* note __, at 112.

⁵¹³ *Id.* at 112.

packaged comparisons?⁵¹⁴

Some argue that the government should focus its energies less on presenting packaged information and more on publishing “reusable data.”⁵¹⁵ The idea, inspired by the engineering principle that separates data from interaction, is that agencies should worry less about designing user-friendly web sites, and more about releasing raw data for nongovernment users.⁵¹⁶ 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.⁵¹⁷ They call this new role for agencies an “invisible hand,” enabling a “marketplace of engineering ideas.”⁵¹⁸ 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).⁵¹⁹

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.⁵²⁰ Perhaps the relevant question, then, is not what policymakers think users *need* to know, but what users *want* to know.⁵²¹ Disclosure policies that consider what information users want, and what they can comprehend, tend to be more successful over time.⁵²²

Moreover, perhaps consumers do not necessarily need more data, but more advice.⁵²³ The opposite of making data “bigger,” of course, is making data “smaller”—usually by simplifying, tailoring, and targeting the information to

⁵¹⁴ 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, *REPUBLIC LOST*, *supra* note __, at 257-58.

⁵¹⁵ David Robinson, Harlan Yu, William P. Zeller, & Edward W. Felten, *Government Data and the Invisible Hand*, 11 *YALE J.L. & TECH.* 160, 160 (2009).

⁵¹⁶ Robinson *et al.*, *supra* note __, at 161.

⁵¹⁷ Robinson *et al.*, *supra* note __, at 161, 169.

⁵¹⁸ Robinson *et al.*, *supra* note __, at 161.

⁵¹⁹ Robinson *et al.*, *supra* note __, 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 __, at 926-27.

⁵²⁰ See, e.g., Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1344; Fenster, *supra* note __, at 942.

⁵²¹ Ben-Shahar & Schneider, *supra* note __, at 746.

⁵²² FUNG ET AL., *supra* note __, at 11.

⁵²³ Ben-Shahar & Schneider, *supra* note __, at 746.

make it easier to process.⁵²⁴ Thus, rating systems and other information made available at the point of purchase could be particularly useful for consumers.⁵²⁵ Mere publication on a government web site 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.⁵²⁶ 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.⁵²⁷ Restaurant hygiene grades, for example, are thus 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.⁵²⁸

Despite the current fascination with “big data,” then, many also appeal for simplification. Agencies are thus thinking more carefully today about ensuring the “utility” of data, perhaps owing in part to the IQA.⁵²⁹ 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.⁵³⁰ 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.⁵³¹ 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.⁵³² For example, there

⁵²⁴ Madison, *supra* note __, 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).

⁵²⁵ FUNG ET AL., *supra* note __, at 57.

⁵²⁶ FUNG ET AL., *supra* note __, at 65-74.

⁵²⁷ FUNG ET AL., *supra* note __, at 57.

⁵²⁸ FUNG ET AL., *supra* note __, at 65.

⁵²⁹ The IQA required the OMB to issue government-wide guidelines for “ensuring and maximizing the quality, objectivity, utility, and integrity of information ... disseminated by the government.” 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.

⁵³⁰ Ben-Shahar & Schneider, *supra* note __, at 675.

⁵³¹ Ben-Shahar & Schneider, *supra* note __, 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)).

⁵³² Archong Fung *et al.*, *Full Disclosure: The Perils and Promise of Transparency* (2007).

is evidence that simple letter grades for restaurant sanitation (from “A” to “C”) have lead to cleaner restaurants in Los Angeles County.⁵³³ 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.⁵³⁴ 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.⁵³⁵ Thus, simple, comprehensible, and easily accessible ratings not only allowed consumers to vote with their wallets, but also encouraged restaurants to compete based on cleanliness—undoubtedly the underlying motivation of the letter grade system.⁵³⁶

For disclosure policies to succeed on multiple levels, then, they must affect not only the decisionmaking of consumers and regulatory beneficiaries, but also the decisionmaking of the discloser—the regulated party.⁵³⁷ Thus, effective disclosure systems become “doubly embedded.”⁵³⁸ The way disclosure policies affect discloser behavior is intuitive—by affecting their profits, market share, and/or reputation.⁵³⁹ 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 themselves are based on complex engineering standards and test results.⁵⁴⁰ Ratings and grades also combat the problem of overdisclosure. Scholars that have evaluated the effectiveness of mandatory disclosure regimes

⁵³³ Ben-Shahar & Schneider, *supra* note __, at 743; Ginger Zhe Jin & Phillip Leslie, *The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards*, 118 Q.J. ECON. (2003) (studying sanitation grades in Los Angeles County).

⁵³⁴ FUNG ET AL., *supra* note __, at 50 (citing studies).

⁵³⁵ FUNG ET AL., *supra* note __, at 50 (citing studies). 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).

⁵³⁶ FUNG ET AL., *supra* note __, at 50-51.

⁵³⁷ FUNG ET AL., *supra* note __, at 65-74.

⁵³⁸ FUNG ET AL., *supra* note __, at 65.

⁵³⁹ FUNG ET AL., *supra* note __, at 66.

⁵⁴⁰ FUNG ET AL., *supra* note __, at 59-61.

sometimes observe that parties can “overdisclose” information to try to “overwhelm and distract” the intended audience.⁵⁴¹ 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.⁵⁴² Unlike dirty restaurants or unsafe cars, it is hard for most people to understand the significance of the terror threat changing from yellow to orange, and more importantly, how to act on that signal.⁵⁴³ 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.

3. Intermediaries and Collaborative Data

Designing databases for use by third party information intermediaries is compelling for several reasons.⁵⁴⁴ Publicly-minded watchdogs like Pro Publica, the Sunlight Foundation, and the Center for Effective Government (formerly OMB Watch) can serve a translational role, sifting large amounts of data into more understandable bits.⁵⁴⁵ Although these organizations focus on government transparency, they also can (and do in fact) help extract and translate information about regulated parties. 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 web sites.⁵⁴⁶ Thus, even raw data sets can be 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

⁵⁴¹ Ben-Shahar & Schneider, *supra* note __, at 700; Lauren E. Willis, *Decisionmaking and the Limits of Disclosure: The Problem of Predatory Lending: Price*, 65 MD. L. REV. 707, 790 (2006).

⁵⁴² FUNG ET AL., *supra* note __, at 61.

⁵⁴³ FUNG ET AL., *supra* note __, at 61.

⁵⁴⁴ For an early discussion of the use of data intermediaries, see Sage, *Regulating Through Information*, *supra* note __, at 1737-41.

⁵⁴⁵ See, e.g., Shkabatur, *supra* note __, at 118.

⁵⁴⁶ FUNG ET AL., *supra* note __, at 62; Goodguide, Scorecard.org: The Pollution Information Site, at <http://scorecard.goodguide.com>; The Right-to-Know Network, at <http://www.rtknet.org>;

provided comprehensible content.”⁵⁴⁷

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). But 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.⁵⁴⁸ 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.⁵⁴⁹ Still, the fact that individuals like Joshua Tauberer (who created Govtrack.us in his spare time)⁵⁵⁰ and Carl Malamud (who painstakingly made SEC data available online),⁵⁵¹ demonstrate that the barriers are far from insurmountable. Moreover, these extraordinary individual efforts inspired the government to publish the data in open formats.⁵⁵²

Not surprisingly, 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.⁵⁵³ 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 companies’ improving safety data, but also the number of jobs they created and taxes they paid by ZIP code.⁵⁵⁴

As such, we might be experiencing a major shift in the government’s informational role from controller to facilitator.⁵⁵⁵ 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 non-profit sectors to maximize its uses.⁵⁵⁶ Data users might also become contributors, as in the case of consumers who report food poisoning

⁵⁴⁷ FUNG ET AL., *supra* note __, at xiv.

⁵⁴⁸ Robinson *et al.*, *supra* note __, at 173.

⁵⁴⁹ Ben-Shahar & Schneider, *supra* note __, at 731-32.

⁵⁵⁰ Robinson *et al.*, *supra* note __, at 165-66, 171.

⁵⁵¹ Robinson *et al.*, *supra* note __, at 166, 171.

⁵⁵² Robinson *et al.*, *supra* note __, at 171.

⁵⁵³ FUNG ET AL., *supra* note __, at 153.

⁵⁵⁴ FUNG ET AL., *supra* note __, at 158-60.

⁵⁵⁵ Indeed, Sage long ago called attention to the government’s role as a data facilitator. Sage, *Regulating Through Information*, *supra* note __, at 1712.

⁵⁵⁶ See, e.g., FUNG ET AL., *supra* note __, at 166.

from restaurants and thus supplement relatively infrequent restaurant inspections.⁵⁵⁷ Just as the CFPB aspires, then, 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.⁵⁵⁸ Such disclosures are mandated by government but are also integrated by disclosers because the information is useful to consumers or users.⁵⁵⁹

Another factor that improves success is whether the information varies between disclosers, such that disclosers have an incentive to race to the top.⁵⁶⁰ 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.⁵⁶¹ Thus, data can serve as an important differentiator between competitors, and they might spend their own resources publicizing differences.⁵⁶² 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 certain data, as in the case of electronic health records (EHRs). Through various pieces of legislation,⁵⁶³ 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.⁵⁶⁴

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.⁵⁶⁵ The goal is to put open, machine-

⁵⁵⁷ FUNG ET AL., *supra* note __, at 166.

⁵⁵⁸ Craswell, *supra* note __, at 369.

⁵⁵⁹ Craswell, *supra* note __, at 369-70.

⁵⁶⁰ Craswell, *supra* note __, at 371.

⁵⁶¹ Craswell, *supra* note __, at 371.

⁵⁶² Craswell, *supra* note __, at 370-71.

⁵⁶³ Most notably the HITECH Act (Health Information Technology for Economic and Clinical Health), Pub. L. No. 111-5, 123 Stat. 226 (2009) (codified as amended in scattered sections of 42 U.S.C.).

⁵⁶⁴ Madison, *supra* note __, at 1618; David Blumenthal & Marilyn Tavenner, *The “Meaningful Use” Regulation for Electronic Health Records*, 363 NEW ENG. J. MED. 501, 501 (2010).

⁵⁶⁵ HealthData.gov, Datasets, <http://www.healthdata.gov/dataset?f%5B0%5D=type%3Adataset> (sorted by “Publisher”).

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.⁵⁶⁶ 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.⁵⁶⁷

The next generation of disclosure will thus be more collaborative, in the sense that various sectors will both contribute to and use the data.⁵⁶⁸ Indeed, there seems to be wide agreement that the government should not have a monopoly on generating data,⁵⁶⁹ 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 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 also blind or 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 deter unwanted behavior.

⁵⁶⁶ HealthData.gov, About, <http://www.healthdata.gov/content/about>.

⁵⁶⁷ Madison, *supra* note __, at 1620-21; Sage, *Regulating Through Information*, *supra* note __, at 1771 (discussing the public good nature of information in the health industry).

⁵⁶⁸ FUNG ET AL., *supra* note __, at 153.

⁵⁶⁹ See, e.g., Esty, *supra* note __, at 199-200.