

REGULATION BY DATABASE

*Nathan Cortez**

- INTRODUCTION 2
- I. FROM OPEN GOVERNMENT TO OPEN INDUSTRY 5
- II. AGENCY DATABASES 12
 - A. The FEC’s Campaign Finance Data 13
 - B. The EPA’s Toxic Release Data 16
 - C. The CFPB’s Consumer Complaint Data 18
 - D. The CPSC’s Product Safety Data 21
 - E. Medicare’s Quality Data 23
 - F. The FDA’s Adverse Event Data 28
- III. DISCLOSURE AND ITS ASPIRATIONS..... 32
 - A. Democratic Ideals 32
 - B. Market Ideals 33
 - C. Regulatory Ideals 34
- IV. DISCLOSURE AND ITS DISCONTENTS 37
- V. DESIGNING FOR OPTIMAL DISCLOSURE 41
 - A. Inputs 42
 - B. Outputs..... 46
 - C. Procedural Safeguards..... 49
- CONCLUSION 53

* Adelfa Botello Callejo Endowed Professor of Law in Leadership & Latino Studies, Associate Dean for Research, SMU Dedman School of Law. J.D. Stanford, B.A. University of Pennsylvania. I thank Elise Legrose and Victoria Vish for characteristically excellent research assistance. This Article builds on my work for the Administrative Conference of the United States (ACUS), which is considering recommendations to address agency use of searchable online databases. *See* ACUS, Agency Information Dissemination in the Internet Era, <https://www.acus.gov/research-projects/agency-information-dissemination-internet-era>; [Cite proposed recommendation once published online.]

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 is publishing thousands upon thousands of searchable online databases, spread among hundreds of federal publishers. The site Data.gov includes 194,770 unique data sets, including roughly 158,000 from the federal government, with thousands more from sub-federal and nongovernmental sources.³ Some consequential portion of these includes information about private parties that is negative, adverse, or unflattering in some way. For example, users can search for consumer products that may have caused injuries; for drugs that may have caused unwanted 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 contributed to federal campaigns; or for local facilities that are discharging toxic chemicals, and in what amounts.

These data are often published not just to promote government transparency or inform the public, but to pursue regulatory aims—to change the underlying behavior being reported. Such “regulation by revelation”⁴ is not entirely new, of course—its lineage stretches back decades. By now, disclosure has been so frequently used as a tool to encourage or 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,⁵ fuel economy,⁶ hospital quality,⁷ mortgages,⁸ securities,⁹ sex offenders,¹⁰ tire safety,¹¹ toxic

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, http://catalog.data.gov/dataset#sec-organization_type (last visited Apr. 1, 2016) (see “Organization Types” in the left sidebar). Data.gov currently includes data from 77 federal agencies and subagencies, and 495 different “publishers.” See Data.gov, Federal Agency Participation, <http://www.data.gov/metrics> (last visited Apr. 1, 2016).

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

⁵ Nutritional Labeling Education Act of 1990, _____.

⁶ U.S. Department of Energy, Download Fuel Economy Data, [Fueleconomy.gov](http://www.fueleconomy.gov/feg/download.shtml), <http://www.fueleconomy.gov/feg/download.shtml> (providing data back to 1978).

⁷ See Part II.E, *infra*.

⁸ Frederick Schauer, *Transparency in Three Dimensions*, 2011 U. ILL. L. REV. 1339, 1341.

⁹ 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,

pollution,¹² and workplace chemical exposure,¹³ among many many other types of conduct. Disclosure has even become a preferred method of regulation internationally.¹⁴

Though policymakers have relied on disclosure-based regulation for decades,¹⁵ in recent years it seems to have evolved from peculiarity to regularity. And, perhaps more importantly, it has evolved in other ways. For example, sometimes the real goal of disclosure is to persuade, rather than inform.¹⁶ 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? Second, disclosure-based regulation is now much less static (aimed narrowly at helping potential users of disclosed information to make better decisions) and more dynamic (aimed more broadly at trying to change the disclosers' underlying behavior).¹⁷ Frequently, the real party being targeted by mandatory disclosure is not the consumer, but the discloser,¹⁸ with the Brandeisian logic that shining a light on undesired behavior will deter it, or at least make it more costly.

Disclosure, then, seems to be in bloom. And this bloom is reflected in the recent proliferation of agency databases, which derives from decades of federal policies promoting government transparency, particularly recent policies pushing the federal government to publish more information online, as detailed in Part I. Online publication has reached a crescendo with the Obama Administration, which published an Open Government Memorandum on the President's first day in office,¹⁹ and has since promoted scores of other transparency projects, including the sites FOIA.gov and Data.gov.

This groundwork is enabling the ongoing swell of government data initiatives, as well as innovative nongovernmental uses of this data. The many thousands of government data sets beget perhaps thousands of third party web sites, mobile applications, and other information products that rely on government data. Perhaps the best example is the government's decision in the 1980s to publish Global Positioning System (GPS) data for civilian use, which has inspired a wave

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

¹⁰ Sex Offender Registration and Notification Act, Pub. L. No. ____ (codified at ____).

¹¹ 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.).

¹² Emergency Planning and Community Right-to-Know Act of 1986, Pub. L. No. _____.

¹³ Occupational Safety and Health Administration, Hazard Communication Standard, 29 C.F.R. § 1910.1200 (requiring material safety data sheets in the workplace).

¹⁴ 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).

¹⁵ See, e.g., Howard Beales, Richard Craswell, & Steven Salop, *The Efficient Regulation of Consumer Information*, 24 J.L. & ECON. 491, 491 (1981).

¹⁶ Ben-Shahar & Schneider, *supra* note __, at 744.

¹⁷ Richard Craswell, Static Versus Dynamic Disclosures, and How Not to Judge Their Success or Failure, 88 WASH. L. REV. 333, 339 (2013).

¹⁸ Ben-Shahar & Schneider, *supra* note __, at 744.

¹⁹ Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government (Jan. 21, 2009), 74 Fed. Reg. 4685 (Jan. 21, 2009).

of innovation that incorporates geospatial, location data,²⁰ such as navigation and restaurant apps. Indeed, when President Obama announced an Open Data Policy, he emphasized that “This kind of innovation and ingenuity has the potential to transform the way we do almost everything.”²¹ Lest the reader discount this as hyperbole, disclosure policies often have seemingly limitless ambitious—many are justified as promoting “autonomy, dignity, civility, community, citizenship, economic growth, and a variety of other virtues.”²² Part II examines these aspirations in light of emerging evidence.

As many wonder, “Who could oppose providing more information to the public?”²³ But as ubiquitous and trendy as disclosure has become, inevitably, criticisms have emerged from scholars who doubt that it is “an unalloyed good.”²⁴ I evaluate the potential harms and shortcomings of disclosure in detail in Part IV. 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.²⁵

Indeed, as Schauer observes, policymakers rarely question whether the burdens of disclosure outweigh its purported benefits.²⁶ One goal of this Article, then, is to help correct this asymmetry, particularly with regard to the growing popularity of one species of disclosure—the searchable online database.

To design an appropriate framework for disclosure, we must understand 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.”²⁷ Similarly, “transparency” can be recharacterized as “adverse transparency” if the information is unflattering or harmful to the subject in some way.²⁸ How, then, can policymakers exercise this

²⁰ 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>.

²¹ 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>.

²² Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 U. PA. L. REV. 647, 734 (2011).

²³ 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.

²⁴ 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).

²⁵ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1342.

²⁶ Ben-Shahar & Schneider, *supra* note ___, at 683 (“[L]awmakers rarely inquire into the effectiveness or burden of disclosure.”).

²⁷ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1347.

²⁸ Sarah Taylor Roller, Raqiyyah R. Pippins, & Jennifer W. Ngai, *FDA’s Expanding Postmarket*

power responsibly? If we value disclosure and accept it as a baseline, how do we manage it?

Part V offers thoughts on how policymakers can design databases for more optimal disclosure, focusing on the inputs and outputs of published data, and procedural safeguards that might help ensure the quality and reliability of data. I envision a decidedly modern role for the government as a “data steward,”²⁹ rather than merely as a source and 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, government agencies can help ensure that data collection practices are fair, that data sets 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. I also propose a series of pre- and post-publication procedures that will help ensure the quality and reliability of government data, building on my recent work for the Administrative Conference of the United States (ACUS), which is currently considering a narrower set of recommendations on the issue.³² We might refer to these recommendations collectively as “Good Government Data Practices.”

As the Article begins, consider the stakes. Successful disclosure policies can help ensure healthy markets, empower consumers, inform citizens, and can even alter the underlying conduct of disclosers. Failed disclosure 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 Thus, as we increasingly rely on agency databases to achieve regulatory ends, both the means and ends require more critical examination.

I. FROM OPEN GOVERNMENT TO OPEN INDUSTRY

Modern efforts to publish the activities of the regulated derive from very old efforts to publish the activities of regulators themselves. For decades, perhaps even centuries,³⁴ we have pushed the federal government to be more open and transparent. Indeed, this long arc toward government “transparency” is a defining hallmark of American administrative law,³⁵ sitting comfortably “among

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).

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, Agency Information Dissemination in the Internet Era, <https://www.acus.gov/research-projects/agency-information-dissemination-internet-era>; [Cite proposed recommendation once published online.]

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

³⁴ 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).

³⁵ William Funk, *Public Participation and Transparency in Administrative Law—Three Examples as an Object Lesson*, 61 *ADMIN. L. REV.* 171 (2009).

the pantheon of great political virtues.”³⁶

But today’s transparency arrived only after decades of reform. During the New Deal, amid the creation of new executive agencies and a flood of new regulations, agencies published regulations on their own, if at all. One of the earliest and most significant efforts to address the lack of federal agency transparency was the Federal Register Act of 1935,³⁷ which created the now-familiar daily gazette of executive documents, including proposed and final rules and other public notices. 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.”³⁸ The Act was motivated in part by the famous “hot oil” case, *Panama Refining*,³⁹ the first Supreme Court decision to invalidate an act of Congress as an excessive delegation of authority.⁴⁰ The “hot oil” law was part of the National Industrial Recovery Act, the flagship New Deal bill passed during the Great Depression.⁴¹ The Act authorized President Roosevelt (who delegated his authority to the Secretary of the Interior), to limit oil production and stabilize prices during ongoing discovery of vast new oil fields in Texas. During the litigation, “the government was embarrassed to admit that a reexamination of the relevant documents (which at the time were not publicly available) had revealed that the Secretary had inadvertently revoked the relevant regulation before the lawsuit had been filed.”⁴² Just weeks before oral argument, 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.⁴³ The “furor” over the case⁴⁴ 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.”⁴⁵ As the federal government bursted 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.”⁴⁶ Thus, before 1935, agencies did not even have to

³⁶ Mark Fenster, *The Opacity of Transparency*, 91 IOWA L. REV. 885, 888 (2006).

³⁷ Pub. L. No. 74-220, ch. 417, 49 Stat. 500 (1935).

³⁸ 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>.

³⁹ 293 U.S. 388 (1935).

⁴⁰ [CITE *Schechter Poultry*].

⁴¹ [CITE *NIRA*].

⁴² Peter Strauss et al., *Gellhorn and Byse’s Administrative Law: Cases and Comments* 446 (11th ed. 2011).

⁴³ 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.

⁴⁴ *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)).

⁴⁵ *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.”).

⁴⁶ 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.

publish the regulations they imposed.⁴⁷

Equally important, the Federal Register Act of 1935 called for the government to publish all federal regulations. Specifically, the 1935 Act called for agencies to compile all existing documents having legal effect prior to publication of the first Federal Register in 1936.⁴⁸ In 1937, Congress amended the Act to require not a “compilation” but a “codification of all regulations every five years.”⁴⁹ A Codification Board recommended organizing the compilation into 50 subjects or titles, similar to the U.S. Code.⁵⁰ Later in 1937, President Roosevelt agreed to the proposed organization, authorizing the government to publish a new *Code of Federal Regulations* (C.F.R.).⁵¹ The Government Printing Office (GPO) thus published the first edition of the C.F.R. in 1939, although precursors were published the previous year in anticipation of the new system.⁵²

Publication requirements were further embedded in American law by the Administrative Procedure Act (APA) of 1946.⁵³ The APA was a response, in part, to complaints from industry that administrative agencies during the New Deal were excessively opaque and insular, particularly towards the private interests most affected by regulation.⁵⁴ 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.⁵⁵ The new procedures creating “notice and comment” rulemaking, in particular, have since become a hallmark of citizen participation in government.⁵⁶ 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.⁵⁷

However, many viewed APA section 3 as more a tool of *withholding* information than disclosing it.⁵⁸ 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 from individual agencies.⁵⁹

Moreover, the APA’s disclosure provisions were largely designed to give notice to

⁴⁷ 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).

⁴⁸ Federal Register Act § 11, 49 Stat. at 503.

⁴⁹ Pub. L. No. 75-158, Act of June 19, 1937, ch. 369, 50 Stat. 304-305.

⁵⁰ McKinney, *supra* note ___, at 10.

⁵¹ Office of the Federal Register, *supra* note ___, at 4.

⁵² *Id.* at 4.

⁵³ Pub. L. No. 79-404, 60 Stat. 237 (1946) (codified as amended in scattered sections of 5 U.S.C.).

⁵⁴ William Funk, Public Participation and Transparency in Administrative Law – Three Examples as an Object Lesson, 61 ADMIN. L. REV. 171, 172-73, 178 (2009).

⁵⁵ APA § 3, 60 Stat. 237 (1946).

⁵⁶ 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).

⁵⁷ 5 U.S.C. § 552(a)(1).

⁵⁸ [CITE].

⁵⁹ STRAUSS ET AL., *supra* note ___, at 451.

those whose legal rights were directly affected by the agency action—“almost invariably businesses.”⁶⁰ 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.⁶¹ As Bill Funk observes, the APA addresses participation in rulemaking to “interested persons,” meaning those with a “direct and palpable interest,”⁶² and required that public records be made available to persons “properly and directly concerned,” rather than the general public.⁶³ Thus, the APA’s disclosure provisions were aimed to inform regulated parties—not shine a light on them, as is the focus of modern efforts.

Thus, if the APA is viewed as a reaction to agency hostility towards private business, then the 1960s and 1970s could be viewed as a movement toward recognizing the public interest on equal footing.⁶⁴ Disclosure and transparency law, as developed by courts and Congress, reflect this general pattern.

The modern open government movement really began in 1967, when Congress passed the Freedom of Information Act (FOIA),⁶⁵ 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.⁶⁶ President Lyndon Johnson, who signed the bill reluctantly and only under pressure from the press corps,⁶⁷ 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.”⁶⁸ Half a century later, modern scholars still acknowledge FOIA’s importance to our democratic government.⁶⁹ In requiring the government to make its records available upon request, unless specifically exempted, FOIA created a “strong presumption in favor of disclosure.”⁷⁰ 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.⁷¹

⁶⁰ Funk, *supra* note __, at 173.

⁶¹ Funk, *supra* note __, at 173; FINAL REPORT OF THE ATTORNEY GENERAL’S COMMITTEE ON ADMINISTRATIVE PROCEDURE 2 (1941), *available at* <http://www.law.fsu.edu/library/admin/1941report.html>.

⁶² Funk, *supra* note __, at 174.

⁶³ 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)).

⁶⁴ Funk, *supra* note __, at 178-80.

⁶⁵ 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.

⁶⁶ See, e.g., 112 CONG. REC. H13007 (daily ed. Jun. 20, 1966) (statement of Rep. John Moss).

⁶⁷ Vladeck, *supra* note __, at 1798.

⁶⁸ 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>.

⁶⁹ 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.”).

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

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

But given its lofty goal of pursuing democracy through transparency and accountability, FOIA has been criticized for falling short of these ideals.⁷² A major complaint is that FOIA produces transparency only by request—it imposes few affirmative disclosure obligations on agencies and instead sometimes requires relatively sophisticated private intermediaries with sufficient “time, money, and expertise” to “press a recalcitrant administration for disclosure.”⁷³ Scholars have also criticized FOIA for being too malleable, particularly by administrations (the George W. Bush administration is frequently cited) that construe its disclosure requirements narrowly and exemptions broadly.⁷⁴ As such, FOIA represents more passive disclosure by agencies, contrasted against actively disclosing or even publicizing the information they hold.⁷⁵ Moreover, the basic premise of FOIA’s “request-and-wait-for-a-response approach” is seen as obsolete in the current Internet era.⁷⁶

Following FOIA was the 1976 Government in the Sunshine Act, a modest addition to the transparency movement, which required meetings of multi-member agencies like the FCC, FTC, and SEC, be open to the public, unless a specific exemption applied.⁷⁷ Because the Act only covered multi-member commissions, it omitted the hundreds of important Cabinet departments or sub-departments, thus covering only 5-10% of all federal agencies.⁷⁸ Moreover, notwithstanding its modest goals, many still view the Sunshine Act as a failure, perhaps even leading to “decreased transparency.”⁷⁹

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 1995,⁸⁰ the Electronic Freedom of Information Act of 1996 (sometimes called “E-FOIA”),⁸¹ and the Government Paperwork Elimination Act of 1998⁸² directed federal agencies to use the Internet to publish more information online and to “improve the productivity, efficiency, and effectiveness of Federal programs.”⁸³ In particular, E-FOIA required agencies to publish online their final opinions and orders, as well as records likely to be requested.⁸⁴ Thus spawned the brief surge in agency online “reading rooms.”

⁷² 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.)).

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

⁷⁴ 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 ____.

⁷⁵ Cortez, *supra* note ___, at ____.

⁷⁶ Vladeck, *supra* note ___, at 1792-93.

⁷⁷ Pub. L. No. 94-409, 90 Stat. 1241 (1976) (codified as amended at 5 U.S.C. § 552b).

⁷⁸ Funk, *supra* note ___, at 188 (citing KENNETH CULP DAVIS, I ADMINISTRATIVE LAW TREATISE 441, 446 (2d ed. 1978)).

⁷⁹ Funk, *supra* note ___, at 197.

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

⁸¹ Pub. L. No. 101-231, 110 Stat. 347 (codified at 5 U.S.C. § 552).

⁸² Pub. L. No. 105-277, div. C, tit. XVII, § 1702, 112 Stat. 2681 (codified at 44 U.S.C. § 3504).

⁸³ Pub. L. No. 104-13, § 2, 109 Stat. 163, 167 (codified at 44 U.S.C. § 3504(h)(5)).

⁸⁴ 5 U.S.C. § 552(a)(2)(D).

In the 2000s, as federal agencies built sprawling web sites,⁸⁵ 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.⁸⁶ The Act also created the Office of Electronic Government within OMB, headed by a Chief Information Officer would coordinate with agencies through a Council.⁸⁷ 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.⁸⁸ 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.⁸⁹

Moreover, during this decade Congress also passed laws requiring more transparency in federal spending, directing the OMB to publish the details of federal grants, loans, and contracts online.⁹⁰ Today, the public can search databases on the federal web site USAspending.gov to view entities that have received money from the federal government,⁹¹ or more specific sites like Recovery.gov to see how the federal government has spent money from the 2009 economic stimulus package.⁹² The former site includes a searchable database with several categories of information, including the name and location of the entity receiving the federal money (including unique numerical identifiers), the amount received, the type of transaction, the funding body, the purpose of the funding, and other information.⁹³ As Vladeck 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.”⁹⁴

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).⁹⁵ The Act required the OMB to issue government-wide guidelines for

⁸⁵ [JIM O'REILLY].

⁸⁶ Pub. L. No. 107-347, 116 Stat. 2899 (codified as amended in scattered sections of the U.S. Code).

⁸⁷ [CITE].

⁸⁸ E-Government Act of 2002, § 101(a), 116 Stat. 2902 (defining “electronic government”).

⁸⁹ Open Government Act of 2007, Pub. L. No. 110-175, 121 Stat. 2524 (codified in scattered sections of 5 U.S.C. § 552).

⁹⁰ 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).

⁹¹ See USAspending.gov, at <http://www.usaspending.gov>.

⁹² 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>.

⁹³ Pub. L. No. 109-282 § 2(b)(1), 120 Stat. at 1187; see also USAspending.gov at <http://www.usaspending.gov>.

⁹⁴ Vladeck, *supra* note ___, at 1829-30 (citing About FedSpending.org, at <http://www.fedspending.org/aboutthissite.php>).

⁹⁵ Treasury and General Government Appropriations Act for Fiscal Year 2001 § 515, Pub. L. No.

“ensuring and maximizing the quality, objectivity, utility, and integrity of information ... disseminated by the government.”⁹⁶ 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.⁹⁷ In 2002, the OMB finalized guidelines implementing the Act,⁹⁸ followed by agencies issuing their own guidelines.⁹⁹

The IQA and OMB guidelines would seem well designed to regulate the quality of information posted in agency databases. Yet the Act’s application to databases is unclear, and might vary by database. The broad wording of the Act states that the OMB guidelines should apply to agency “dissemination of public information, regardless of the form or format.”¹⁰⁰ And the OMB guidelines define “information” as “any communication or representation of knowledge such as facts or data, in any medium or form,”¹⁰¹ including “information that an agency disseminates from a web page.”¹⁰² 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.”¹⁰³ The guidelines also exempt “adjudicative processes.”¹⁰⁴ These exemptions might thus exclude important agency databases, such as the CFPB’s Consumer Complaint Database.

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

In December 2009, the OMB published the Open Government Directive,¹⁰⁸ following on President Obama’s Open Government Memorandum. The directive

106-554, 114 Stat. 2763, 2763A-153-54 (2001); 44 U.S.C. § 3516.

⁹⁶ *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.

⁹⁷ *Id.*

⁹⁸ 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).

⁹⁹ OMB, Agency Information Quality Guidelines, https://www.whitehouse.gov/omb/inforeg_agency_info_quality_links/ (last visited June 15, 2015).

¹⁰⁰ 44 U.S.C. § 3504(d)(1).

¹⁰¹ 67 Fed. Reg. at 8458-59.

¹⁰² *Id.* at 8460.

¹⁰³ 67 Fed. Reg. at 377.

¹⁰⁴ *Id.*

¹⁰⁵ Memorandum for the Heads of Executive Departments and Agencies: Transparency and Open Government (Jan. 21, 2009), 74 Fed. Reg. 4685 (Jan. 21, 2009). As Shkabatur notes, “[d]ozens of other countries” have followed this example. 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>).

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

¹⁰⁷ Gellhorn & Byse’s Administrative Law, at 441.

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

urged agencies to “take prompt steps to expand access to information by making it available online in open formats.”¹⁰⁹ The directive required each agency, within 45 days, to “identify and publish online in an open format at least three high-value data sets ... on Data.gov” that had previously not been available.¹¹⁰ Within 60 days, each agency was to create an Open Government web page. Today, 78 different federal agencies and subagencies have posted 194,126 datasets on Data.gov.¹¹¹

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

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

II. AGENCY DATABASES

What has this long history of disclosure initiatives produced? Thousands of federal agency web sites, and hundreds of thousands of agency databases (as of March 2016, over 194,000).¹¹⁷ As Ben-Shahar and Schneider observe, writing about disclosure is difficult because it is so very common.¹¹⁸ 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.¹¹⁹ Nevertheless, as becomes quickly obvious by early

¹⁰⁹ *Id.*

¹¹⁰ *Id.* at 7.

¹¹¹ Federal Agency Participation, Data.gov, at <http://www.data.gov/metric> (last visited March 22, 2016).

¹¹² FOIA.gov, U.S. Department of Justice, at <http://www.foia.gov>.

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

¹¹⁴ Shkabatur, *supra* note ___, at 100.

¹¹⁵ Also making this observation are FUNG ET AL., *supra* note ___, at xii-xiii. They call this “targeted transparency.”

¹¹⁶ FUNG ET AL., *supra* note ___, at 28.

¹¹⁷ 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).

¹¹⁸ Ben-Shahar & Schneider, *supra* note ___, at 652.

¹¹⁹ *Id.* at 652.

efforts to survey agency databases,¹²⁰ they are becoming the norm rather than the exception.

Below I survey six 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 contributions and spending, but also to pursue deeper goals of deterring corruption and the appearance of it.¹²¹ Indeed, disclosure has been perhaps the one leg of the campaign finance law tripod to be spared by the Supreme Court. Unlike limits on contributions and expenditures, disclosure requirements have endured repeated First Amendment challenges—from early cases like *Burroughs*¹²² to more aggressive recent cases like *Citizens United* and *McCutcheon*.¹²³ Of course, in the seminal case *Buckley v. Valeo*,¹²⁴ the Court upheld various disclosure requirements introduced in the Federal Election Campaign Act of 1971 (FECA) and the FECA Amendments of 1974.¹²⁵ In the ensuing decades, as various restrictions on campaign contributions and expenditures fell under First Amendment scrutiny, disclosure requirements endured. Thus, disclosure has become, through attrition, the preferred choice for regulating money in politics.¹²⁶ Indeed, recent reform proposals now seek to address lingering problems by using even more disclosure, for example by trying to triangulate FEC data with data from other agencies like the SEC and IRS.¹²⁷

Today, the FEC maintains several searchable online databases on its web site, which are centralized on the FEC's Campaign Finance Disclosure Portal.¹²⁸ Users can search FEC data based on reports required of federal candidates, parties, committees, donors, and lobbyists, among others.¹²⁹ The data are generally searchable by name, date, and location,¹³⁰ and are presented in list, map, and

¹²⁰ Cortez, ACUS Report, Appendix E.

¹²¹ *Buckley v. Valeo*, 424 U.S. 1, 66-68 (1976).

¹²² *Burroughs v. United States*, 290 U.S. 534 (1934) (upholding an indictment under the Federal Corrupt Practices Act).

¹²³ *Citizens United v. FEC*, 558 U.S. 310, 319 (2010); *McCutcheon v. FEC*, 134 S.Ct. 1434, 1442, 1459-60 (2014).

¹²⁴ 424 U.S. 1 (1976).

¹²⁵ 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).

¹²⁶ 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).

¹²⁷ See, e.g., Lucian A. Bebchuk & Robert J. Jackson, Jr., *Shining Light on Corporate Political Spending*, 101 GEO. L.J. 923 (2013).

¹²⁸ FEC, Campaign Finance Disclosure Portal, <http://www.fec.gov/pindex.shtml>.

¹²⁹ 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/>.

¹³⁰ FEC, Searchable Presentations, <http://www.fec.gov/portal/searchable.shtml>.

chart form.¹³¹ 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.¹³² The FEC's Disclosure Data Catalog publishes these datasets in downloadable .CSV, .XML, or .XSD formats.¹³³

Despite their broad scope, FEC databases pay special attention, of course, to 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.¹³⁴

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.¹³⁵ The system includes, for example, a searchable and sortable list of parties required to pay administrative fines for violating reporting requirements.¹³⁶ 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.¹³⁷ The Act requires House and Senate officials to use “computerized systems” with “coding” and “cross-indexing” to “maximize public access to materials filed.”¹³⁸ Congress also requires these reports to be available over the Internet.¹³⁹ 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.¹⁴⁰ The 2007 amendments also linked lobbying information with campaign contribution data reported to the FEC,¹⁴¹ so that users can better track financial ties between lobbyists and public officials.

As with many other disclosure-based regulatory schemes, researchers question how effectively campaign finance disclosures have achieved their stated goals of

¹³¹ See, e.g., FEC, 2016 Presidential Campaign Finance, <http://www.fec.gov/disclosure/pnational.do>.

¹³² FEC, Disclosure Data Catalog, <http://www.fec.gov/data/DataCatalog.do?cf=searchable>.

¹³³ FEC, Disclosure Data Catalog, <http://www.fec.gov/data/DataCatalog.do>.

¹³⁴ See, e.g.,

¹³⁵ FEC, FEC Enforcement Query System (EQS), at <http://eqs.fec.gov/eqs/searcheqs>.

¹³⁶ FEC, Administrative Fines, <http://www.fec.gov/data/AdminFine.do?format=html>.

¹³⁷ Pub. L. No. 104-65, 109 Stat. 691, codified as amended at 2 U.S.C. §§ 1601-07.

¹³⁸ *Id.* at § 6.

¹³⁹ *Id.* at § 6(a)(4), (9).

¹⁴⁰ 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>.

¹⁴¹ Honest Leadership & Open Government Act, *supra* note ___, at § ___.

preventing corruption and the appearance of it.¹⁴² Compliance with FEC disclosure requirements are 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.¹⁴³ 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.¹⁴⁴ Thus, the data “gives the illusion of transparency, but functions instead to obscure the most pertinent financial constituencies in a sea of data.”¹⁴⁵ This design failure means that the intended users—voters, intermediaries, and regulators—must rely on flawed data.¹⁴⁶

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.¹⁴⁷ The preconditions are substantially similar to the conditions prescribed by Fung, Graham, and Weil (which I discuss in more detail in Part V, *infra*).¹⁴⁸ There are numerous articles by campaign finance scholars suggesting specific improvements to the FEC’s data practices, including how it collects, processes, and disseminates data, and enforces compliance.¹⁴⁹ Simple fixes, like assigning unique identifiers to contributors, and allowing the FEC to conduct random audits, might improve the quality of the data greatly.¹⁵⁰

As with other databases, informational intermediaries play an important role in translating and synthesizing campaign finance data. Groups like the Center for Responsive Politics (CRP), the Campaign Finance Institute (CFI), and the Sunlight Foundation rely on FEC data to provide more digestible information to the public.¹⁵¹ 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.¹⁵² Another group called MAPlight.org mashes up the publicly-available voting records of members of Congress with campaign finance data, trying to find correlations.¹⁵³ Maplight gathers data from not only the FEC, but also from the Center for Responsive Politics, which runs OpenSecrets.org, and

¹⁴² Heerwig & Shaw, *supra* note __.

¹⁴³ Heerwig & Shaw, *supra* note __, at 1479-85.

¹⁴⁴ Heerwig & Shaw, *supra* note __, at 1484.

¹⁴⁵ Heerwig & Shaw, *supra* note __, at 1484.

¹⁴⁶ Heerwig & Shaw, *supra* note __, at 1486-89.

¹⁴⁷ Michael J. Malbin & Thomas L. Gais, *The Day After Reform: Sobering Campaign Finance Lessons from the American States* 36 (1998).

¹⁴⁸ Heerwig & Shaw, *supra* note __, at 1475 (citing Fung *et al.*, *supra* note __, at 11).

¹⁴⁹ 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).

¹⁵⁰ Heerwig & Shaw, *supra* note __, at 1494-99.

¹⁵¹ Heerwig & Shaw, *supra* note __ 1487-88.

¹⁵² 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).

¹⁵³ Maplight.org, <http://www.maplight.org>.

FollowTheMoney.org (for California data).¹⁵⁴ 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.¹⁵⁵

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.¹⁵⁶ 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.”¹⁵⁷

As David Vladeck observes, several well-known environmental statutes are predicated on disclosure, such as the Clean Air Act, the Emergency Planning and Community Right-to-Know Act, the Federal Insecticide, Fungicide, and Rodenticide Act, the National Environmental Policy Act, the Safe Drinking Water Act, and the Toxic Substances Control Act.¹⁵⁸ The commonality is that these statutes all “place affirmative duties on federal agencies to make information available to the public.”¹⁵⁹ In combination, these statutes “seem to provide a right of public access to virtually all environmental information in the hands of the federal government.”¹⁶⁰

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.”¹⁶¹ The database became the Toxics Release Inventory (TRI) program. The program requires facilities to report their production and release of roughly 650 dangerous chemicals.¹⁶² The EPA first reported TRI data in 1989, and first published it online in 1998.¹⁶³ 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.¹⁶⁴

¹⁵⁴ Maplight.org, Find Contributions, About or Data (at the bottom of the page), <http://maplight.org/us-congress/contributions>.

¹⁵⁵ Heerwig & Shaw, *supra* note __, at 1490.

¹⁵⁶ 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).

¹⁵⁷ Environmental Law Institute, *The Environmental Forum* 36 (July/August 1998).

¹⁵⁸ 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).

¹⁵⁹ *Id.* at 1788.

¹⁶⁰ *Id.* at 1788.

¹⁶¹ Pub. L. No. 99-499 § 313, codified as amended at 42 U.S.C. §§ 11023(h), (j).

¹⁶² 30 C.F.R. § 372.65; *see also* 42 U.S.C. § 11023 [YEAR].

¹⁶³ EPA, TRI Program Milestones, at

http://www.epa.gov/sites/production/files/documents/tritimeline_0.pdf.

¹⁶⁴ EPA, Toxics Release Inventory (TRI) Program, <http://www.epa.gov/toxics-release-inventory-tri-program>.

The data are presented in colorful chart, graph, and map forms.¹⁶⁵ 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.¹⁶⁶

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.¹⁶⁷ The initial success of the TRI program surprised the EPA and environmental groups, who had toiled for years to regulate toxic pollution.¹⁶⁸ As several scholars note, ten years after initiating the TRI program, the amount of pollution released had dropped by half.¹⁶⁹ In fact, the mere 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.¹⁷⁰ As one of the earliest programs of its kind, the TRI has been applauded by many as the best example of regulation via disclosure.¹⁷¹

Perhaps inspired by its own success, the EPA now publishes hundreds of datasets online. The EPA web site publishes so many datasets that it includes several landing pages that help users search for and navigate the data available.¹⁷² 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.¹⁷³ The EPA currently lists 96 datasets on its web site,¹⁷⁴ with 1,738 listed on Data.gov.¹⁷⁵ The agency even hosts an online discussion forum for data developers.¹⁷⁶

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.¹⁷⁷ They found, in particular, that toxic release data is not embedded in potential users’ decisionmaking, as “[m]ost home buyers,

¹⁶⁵ 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.

¹⁶⁶ EPA, Envirofacts, at <https://www3.epa.gov/enviro/index.html>.

¹⁶⁷ 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).

¹⁶⁸ FUNG ET AL., *supra* note ___, at 29.

¹⁶⁹ FUNG ET AL., *supra* note ___, at 29.

¹⁷⁰ FUNG ET AL., *supra* note ___, at 29.

¹⁷¹ [CITE].

¹⁷² 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>.

¹⁷³ EPA, Enforcement and Compliance History Online (ECHO), at <https://echo.epa.gov>.

¹⁷⁴ EPA, Data – Developer Central, at <https://developer.epa.gov/category/data/>.

¹⁷⁵ 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).

¹⁷⁶ EPA, Data and Developer Forum, at <http://blog.epa.gov/data/>. The forum does not appear to be particularly active.

¹⁷⁷ FUNG ET AL., *supra* note ___, at 76, 79.

renters, job seekers, consumers, and investors do not consider toxic pollution” when making decisions that might be affected by it.¹⁷⁸ The data did succeed in better informing policymakers in Congress and the EPA.¹⁷⁹ And many manufacturers quickly embedded the new data into their decisionmaking, recognizing the reputational and regulatory consequences.¹⁸⁰ 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.¹⁸¹ Scholars have long worried that EPA datasets are “patchy” and “unreliable.”¹⁸²

Moreover, in the last decade, despite its success, the EPA has “drastically scaled back the information made public” under the TRI program.¹⁸³ Prior to 2006, the EPA required facilities to report information regarding any chemical release over 500 pounds.¹⁸⁴ 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.¹⁸⁵ The GAO criticized the EPA’s rule as reducing the “quantity and detail of information” released to communities.¹⁸⁶

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.”¹⁸⁷ 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.¹⁸⁸ Of course, even data optimists believe that there are significant challenges in producing data that is reliable and usable.¹⁸⁹ 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.¹⁹⁰ [EXAMPLES?]

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 particularly interesting. Born of the 2008 financial crisis, the Bureau was “designed in a world of new

¹⁷⁸ FUNG ET AL., *supra* note __, at 85.

¹⁷⁹ FUNG ET AL., *supra* note __, at 86.

¹⁸⁰ FUNG ET AL., *supra* note __, at 86.

¹⁸¹ FUNG ET AL., *supra* note __, at 86 (citing studies).

¹⁸² Esty, *supra* note __, at 156.

¹⁸³ Vladeck, *supra* note __, at 1791.

¹⁸⁴ 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).

¹⁸⁵ *Id.* at 76,937.

¹⁸⁶ 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.

¹⁸⁷ Esty, *supra* note __, at 119.

¹⁸⁸ Esty, *supra* note __, at 119.

¹⁸⁹ Esty, *supra* note __, at 171-74.

¹⁹⁰ Esty, *supra* note __, at 208-09.

technology.”¹⁹¹ The CFPB “is a new agency operating under a new statute and is on the frontier of the open data trend.”¹⁹² The Bureau’s Office of Consumer Response operates a process for consumers to file complaints regarding financial products and services. Complaints, quite notably, are published online by the CFPB in a massive searchable Consumer Complaint Database, which identifies companies by name.¹⁹³

The CFPB web site allows consumers to file complaints for 11 categories of financial products, including mortgages, student loans, and credit cards.¹⁹⁴ Complaints can specify the name of the company, the type of product or service, the type of problem, and the consumer’s zip code, which the Bureau authenticates and sends to the company for a response.¹⁹⁵ The database is searchable and sortable, and includes not only the company’s name, but also its response and whether the response was timely or further disputed by the consumer.¹⁹⁶ Companies select responses via pull-down menu that includes responses such as “Closed with monetary relief,” “Closed with non-monetary relief,” “Incorrect company,” and “In progress.”¹⁹⁷ 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.¹⁹⁸ Each complaint and response is published, but only if they meet several publication criteria.¹⁹⁹

In 2014, the Bureau reversed its position on posting in the database narrative comments by consumers.²⁰⁰ It originally declined to do so, citing consumer privacy and the risk of disclosing consumers’ personal information.²⁰¹ 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.²⁰² As a measure of symmetry, the Bureau proposed to allow

¹⁹¹ Katherine Porter, *The Complaint Conundrum: Thoughts on the CFPB’s Complaint Mechanism*, 7 BROOK. J. CORP. FIN. & COM. L. 57, 86 (2012).

¹⁹² 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).

¹⁹³ CFPB, Consumer Complaint Database, at <http://www.consumerfinance.gov/complaintdatabase/>.

¹⁹⁴ *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).

¹⁹⁵ 77 Fed. Reg. at 37,616-17.

¹⁹⁶ Cortez, ACUS Report, *supra* note ___, at 62-63.

¹⁹⁷ CFPB, Company Portal Manual (Version 2.14) (May 2015), <http://www.cfbjlaw.com/files/Uploads/Documents/Resources/cfpb-company-portal-manual-may-2015.pdf>.

¹⁹⁸ *Id.* at 24.

¹⁹⁹ The Bureau will not publish complaints if they are missing critical information, have been referred to other regulators, are duplicative, would reveal trade secrets, are fraudulent, or identify the wrong company. *Id.* at 26; CFPB, Final Policy Statement: Disclosure of Consumer Complaint Data, 78 Fed. Reg. 21,218, 21,225 (Apr. 10, 2013).

²⁰⁰ 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).

²⁰¹ 77 Fed. Reg. at 37,568.

²⁰² CFPB, Final Policy Statement: Disclosure of Consumer Complaint Narrative Data, 80 Fed.

companies to post their own narrative responses,²⁰³ 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.”²⁰⁴ But these responses are optional; companies need not select one for publication.²⁰⁵ Today, 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.²⁰⁶

The intended users of the Complaint Database are consumers, researchers, the Bureau, other regulators, and even the subjects of the complaints themselves—companies.²⁰⁷ The Bureau and consumers groups emphasize that the primary intended beneficiary is consumers, and that publishing complaint data is a “public service” that can “empower” consumers and help them avoid “bad actors” in these markets.²⁰⁸ The Bureau itself emphasizes that “disclosure is one of the best tools government agencies can use.”²⁰⁹ 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.”²¹⁰ Bureau staff hope that intermediaries develop mobile apps and other information products based on complaint data.²¹¹

The 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.²¹² Public interest research groups (PIRGs) are producing reports for certain financial products, like credit cards and debt collection, based on Bureau data.²¹³ In 2015, the rankings-crazed *U.S. News & World Report* ranked credit cards relying, in part, on data from the Consumer Complaint Database.²¹⁴

Another parallel with other agencies is that the Bureau hopes that the act of publication itself will encourage companies to improve their underlying behavior.²¹⁵ Bureau staff report that complaints have inspired some companies, for example, to address potential problems, such as long customer service phone

Reg. 15,572 (Mar. 24, 2015).

²⁰³ 79 Fed. Reg. at 42,768.

²⁰⁴ Company Portal Manual, *supra* note ___, at 29.

²⁰⁵ 80 Fed. Reg. at 15,583.

²⁰⁶ Cortez, ACUS Report, *supra* note ___, at 65.

²⁰⁷ Cortez, ACUS Report, *supra* note ___, at 65; 76 Fed. Reg. at 76,630-31.

²⁰⁸ 78 Fed. Reg. at 21,220.

²⁰⁹ 78 Fed. Reg. at 21,220.

²¹⁰ Richard Cordray, Director, CFPB, Remarks at the Consumer Response Field Hearing (Mar. 28, 2013), <http://www.consumerfinance.gov/speeches/prepared-remarks-of-director-richard-cordray-at-the-consumer-response-field-hearing/>.

²¹¹ Cortez, ACUS Report, *supra* note ___, at 66.

²¹² 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).

²¹³ 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>.

²¹⁴ 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).

²¹⁵ Cortez, ACUS Report, *supra* note ___, at 66.

trees.²¹⁶ 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.²¹⁷ Consulting firms now advise companies to “turn what they hear from the CFPB’s consumer complaint database into a business advantage.”²¹⁸ Bureau staff also report that some companies have tied executive compensation to how well the company has responded to complaints.²¹⁹ In short, the Bureau sees its role somewhat modestly as a publisher—providing a window into a dialogue between companies and their customers.²²⁰ Once complaints are published, “The marketplace of ideas then does the rest.”²²¹

Of course, the Consumer Complaint Database has not been without controversy, 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.²²² 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 complaints online.²²³

The Bureau has fielded industry complaints about the database in several formats, including public comments filed during notice-and-comment periods, in letters to the Director, and in complaints to the Bureau’s Ombudsman.²²⁴ 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.”²²⁵

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.

D. The CPSC’s Product Safety Data

²¹⁶ Cortez, ACUS Report, *supra* note ___, at 66.

²¹⁷ 76 Fed. Reg. at 76,630 n.9.

²¹⁸ 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>.

²¹⁹ Cortez, ACUS Report, *supra* note ___, at 66.

²²⁰ Cortez, ACUS Report, *supra* note ___, at 66-67.

²²¹ 76 Fed. Reg. at 76,631.

²²² Cortez, ACUS Report, *supra* note ___, at 67-70.

²²³ Cortez, ACUS Report, *supra* note ___, at 67-71.

²²⁴ Cortez, ACUS Report, *supra* note ___, at 75.

²²⁵ 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”).

The CPSC’s database, SaferProducts.gov, enjoys more clear statutory authority and design attention from Congress than most 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.”²²⁶ But in 2008, Congress passed the Consumer Product Safety Improvement Act, requiring the agency to create a searchable online database of product safety incidents.²²⁷ Congress 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.²²⁸ The Act requires each report to describe the product or substance at issue, identify the manufacturer or labeler, and describe the harm reported.²²⁹ 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.”²³⁰ The CPSC launched the site, SaferProducts.gov, in March 2011.²³¹

SaferProducts.gov includes a searchable online database of thousands “reports of harm” for specific identified products.²³² (The site also has a portal for users to report incidents, and a portal for companies to respond to them.²³³) The database is keyword-searchable, with advanced search options that include product name, company or brand name, and the product model.²³⁴ 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.²³⁵

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.”²³⁶ As such, SaferProducts.gov includes a disclaimer that tracks this language almost verbatim.²³⁷

One innovation in SaferProducts.gov, a feature that might be emulated by other

²²⁶ 15 U.S.C. § 2051(b).

²²⁷ Consumer Product Safety Improvement Act of 2008 (CPSIA) § 212, Pub. L. No. 110-314, 122 Stat. 3016 (codified at 15 U.S.C. 2055a).

²²⁸ 15 U.S.C. § 2055a(b)(1)(A).

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

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

²³¹ 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).

²³² 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/>.

²³³ 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>.

²³⁴ CPSC, SaferProducts.gov, Advanced Search, at <http://www.saferproducts.gov/Search/AdvancedSearch.aspx>.

²³⁵ Id.

²³⁶ 15 U.S.C. § 2055a(b)(5).

²³⁷ 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 Consumer Product Safety Information Database on SaferProducts.gov, particularly with respect to information submitted by people outside of CPSC.”).

agency databases,²³⁸ is that the CPSC allows manufacturers to comment on reports and object to inaccuracies. The Act itself 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.²³⁹ Likewise, the CPSC must consider objections that any information in a report is “materially inaccurate,”²⁴⁰ 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.”²⁴¹

Indeed, a series of inaccurate product safety warnings inspired Congress to amend the Consumer Product Safety Act in 1981.²⁴² In a recent case a company anonymously challenged an inaccurate product safety report on SaferProducts.gov [describe]. 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.’”²⁴³ The court sustained the company’s challenge, somewhat remarkably,²⁴⁴ 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.²⁴⁵

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.²⁴⁶

The GAO found, for example, that some users were confused about the purpose of SaferProducts.gov, thinking it was a site that featured safe rather than unsafe products.²⁴⁷ Indeed, the “upbeat” name of the database might have contributed to the misperceptions.²⁴⁸ 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.²⁴⁹

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,

²³⁸ Cortez, ACUS Report, *supra* note ___, at 20-21.

²³⁹ 15 U.S.C. § 2055a(c)(2).

²⁴⁰ 15 U.S.C. § 2055a(c)(4).

²⁴¹ 16 C.F.R. § 1102.26(a)(1).

²⁴² 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).

²⁴³ *Company Doe v. Tenenbaum*, 900 F. Supp. 2d 572, 597 (D. Md. 2012).

²⁴⁴ Cortez, ACUS Report, at Appendix ___ (Table of Federal Cases 197_ - 2014) (surveying similar cases and finding few successful challenges).

²⁴⁵ 900 F. Supp. 2d at 597.

²⁴⁶ CPSC, The CPSC Open Government Plan, at <http://www.cpsc.gov/open>.

²⁴⁷ 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>.

²⁴⁸ *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).

²⁴⁹ *Id.* at 12-14.

and other types of providers, 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 overwhelming.²⁵⁰

To facilitate such choices, CMS operates five searchable databases that compare Medicare providers—Hospital Compare,²⁵¹ Physician Compare,²⁵² Nursing Home Compare,²⁵³ Home Health Compare,²⁵⁴ and Dialysis Facility Compare.²⁵⁵ The search functions all work in roughly the same way: users can search for providers by city, state, or by 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, requiring the user to also search for a physician’s name, specialty, or medical condition to help narrow the results.²⁵⁶

The five databases offer quite distinct quality data. Hospital Compare includes data on over 4,000 hospitals nationwide,²⁵⁷ 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.²⁵⁸ The data displayed is 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.²⁵⁹ The tab titled “Payment & Value of Care” might tantalize health policy wonks, but unfortunately, comparative data often is not displayed directly (“Get Results for This Hospital” when searching for Medicare spending per beneficiary), or meaningfully (“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

²⁵⁰ [CITE HALL & SCHNEIDER ARTICLE; JOST BOOK]

²⁵¹ CMS, Hospital Compare, at <https://www.medicare.gov/hospitalcompare/search.html>.

²⁵² CMS, Physician Compare, at <https://www.medicare.gov/physiciancompare/search.html>.

²⁵³ CMS, Nursing Home Compare, at <https://www.medicare.gov/nursinghomecompare/search.html>.

²⁵⁴ CMS, Home Health Compare, at <https://www.medicare.gov/homehealthcompare/search.html>.

²⁵⁵ CMS, Dialysis Facility Compare, at <https://www.medicare.gov/dialysisfacilitycompare/#search>.

²⁵⁶ CMS, Physician Compare, About the Data, at <https://www.medicare.gov/physiciancompare/staticpages/data/aboutthedata.html>.

²⁵⁷ CMS, What is Hospital Compare?, at <http://www.medicare.gov/hospitalcompare/About/What-Is-HOS.html>.

²⁵⁸ 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.

²⁵⁹ CMS, Hospital Compare, Footnotes, at <https://www.medicare.gov/hospitalcompare/Data/Footnotes.html>.

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.”²⁶⁰ Still, given how much the data is incomplete, it is questionable whether it meets even this goal.

Physician quality data is even more limited. 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 relevant board certifications. But there are few quality metrics. Physician Compare does feature a Physician Quality Reporting System, which asks physicians to report on whether they follow certain best practices.²⁶¹ 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.²⁶² Although there are over 200 reportable quality measures, few are listed on Physician Compare.²⁶³

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 high-quality 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 health care.²⁶⁴

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.²⁶⁵ 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

²⁶⁰ 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).

²⁶¹ For a list of examples, see CMS, Physician Compare, Physician Quality Reporting System (PQRS), at <https://www.medicare.gov/physiciancompare/staticpages/data/pqrs.html>.

²⁶² *Id.*

²⁶³ 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.

²⁶⁴ Madison, *supra* note ___, at 1625.

²⁶⁵ Madison, *supra* note ___, at 1626; CMS, Hospital Compare, at <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html>.

readmission rates, among other many others.²⁶⁶ Today's Compare databases are an amalgam of data from a variety of sources, gradually added like ornaments to a Christmas tree.²⁶⁷

Perhaps the richest source of data is Medicare claims data. CMS processes over a billion Medicare claims each year,²⁶⁸ with each claim including multiple data points, including whom Medicare is paying and for what. Medicare claims data has long been used by academics, government researchers, and providers themselves to better understand the U.S. health care system.²⁶⁹ The growth of "big data" in health care, in fact, roughly parallels and depends on the growth and accessibility of Medicare claims data.²⁷⁰ Although Medicare has released various data to the public for years, CMS was restricted from releasing physician claims data by court order.²⁷¹ In 2014, a year after the order was lifted, CMS released claims data for over 880,000 providers.²⁷² [MORE]

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,²⁷³ and today over 1,300 hospitals report data about infection rates, mortality rates, and other data points in order to boost their Medicare reimbursements,²⁷⁴ 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.²⁷⁵ 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.²⁷⁶ Thus, various data sources are being combined in novel ways.

Nevertheless, scholars have long questioned the utility of performance data and

²⁶⁶ Madison, *supra* note ___, at 1626 (citing CMS, Hospital Compare, at <https://www.cms.gov/medicare/quality-initiatives-patient-assessment-instruments/hospitalqualityinits/hospitalcompare.html>).

²⁶⁷ As with other agencies, the raw data files at a separate site. See CMS, Data.Medicare.gov, at <http://data.medicare.gov>.

²⁶⁸ 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).

²⁶⁹ Madison, *supra* note ___, at 1609-10.

²⁷⁰ Madison, *supra* note ___, at 1610.

²⁷¹ 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).

²⁷² 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>.

²⁷³ 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)).

²⁷⁴ CMS, Medicare Program, 78 Fed. Reg. 50,496, 50,678-80 (Aug. 19, 2013)

²⁷⁵ 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.

²⁷⁶ 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.

report cards in the health industry.²⁷⁷ Patients report that they seldom rely on publicly available data or even more comprehensible report cards of physicians, hospitals, or procedures.²⁷⁸ 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.²⁷⁹ Patients frequently are not aware of the information, do not understand it, or do not use it.²⁸⁰ Indeed, of numerous disclosure regimes analyzed by Fun, Graham, and Weil, they found patient safety disclosures to be among the least effective regimes, due to the complexity of the information, cognitive biases, the likelihood the data would be misinterpreted, and the risk of strategic behavior by providers.²⁸¹ As a result, CMS has tried to “shrink” the voluminous and varied data by translating it into star ratings, using a five-star scale.²⁸² 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 are the studies showing that those being measured—hospitals, physicians, and other providers—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.²⁸³ For example, a study of cardiac surgery report cards in Pennsylvania found that cardiac surgeons responded to the new disclosure requirement by becoming more reluctant to operate on sicker patients.²⁸⁴ 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.²⁸⁵

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.”²⁸⁶ One study found that less than 1% of

²⁷⁷ 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); [OTHERS].

²⁷⁸ Id.

²⁷⁹ 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.

²⁸⁰ Ben-Shahar & Schneider, *supra* note __, at 672, 711-12 (citing studies).

²⁸¹ FUNG ET AL., *supra* note __, at 76-77 (analyzing disclosure requirements in Pennsylvania and New York).

²⁸² See, e.g., CMS, *Nursing Home Compare*, <http://www.medicare.gov/nursinghomecompare/search.html>; Madison, *supra* note __, at 1627.

²⁸³ 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).

²⁸⁴ Eric C. Schneider & Arnold M. Epstein, *Use of Public Performance Reports: A Survey of Patients Undergoing Cardiac Surgery*, 279 (20) JAMA 1638 (1998).

²⁸⁵ FUNG ET AL., *supra* note __, at 89 (citing studies)

²⁸⁶ 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*

patients knew how their hospital or surgeon was rated under mandated ratings systems.²⁸⁷ 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.²⁸⁸ The mistaken star rating was published on the CMS web site, and did not correct it for almost two years (from July 2010 to February 2012).²⁸⁹ 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.²⁹⁰ 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."²⁹¹ 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.²⁹² 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.²⁹³ 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 status or reimbursement status, or some other tangible harm.²⁹⁴ 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.²⁹⁵ [Note appendix of federal cases in ACUS report with similar findings on due process claims.]

Still, quality ratings and disclosure policies remain the rage in health policy.²⁹⁶ Contemporary health policy is littered with patient surveys, outcomes data, star ratings or rankings, and of course federal databases mentioned above that combine many of these data points.²⁹⁷

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,²⁹⁸ Warning Letters,²⁹⁹ recalls,³⁰⁰ and enforcement

note ___, at 1694.

²⁸⁷ Epstein, *Rolling Down the Runway*, *supra* note ___, at 1694.

²⁸⁸ *Bryn Mawr Care v. Sebelius*, 898 F. Supp. 2d 1009, 1011 (N.D. Ill. 2012).

²⁸⁹ *Id.*

²⁹⁰ *Id.* at 1011-12.

²⁹¹ *Id.* at 1012.

²⁹² *Id.* at 1013.

²⁹³ *Id.* at 1014.

²⁹⁴ *Id.* at 1014-15.

²⁹⁵ *Id.* at 1018.

²⁹⁶ [HEALTH AFFAIRS]

²⁹⁷ See, e.g., FUNG ET AL., *supra* note ___, at 160-62 (listing various disclosure systems in health care).

²⁹⁸ FDA, Inspections Database, at <http://www.fda.gov/ICECI/Inspections/ucm222557.htm>.

²⁹⁹ FDA, Inspections, Compliance, Enforcement, and Criminal Investigations: Warning Letters, at <http://www.fda.gov/ICECI/EnforcementActions/WarningLetters/>.

reports.³⁰¹ The FDA also maintains a database of good news—product approvals.³⁰² 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 medication errors and adverse drug events reported to the agency.³⁰³ Adverse event reporting dates back at least 35 years, and perhaps even longer, when the agency received reports by paper.³⁰⁴ Manufacturer reports are required by regulation,³⁰⁵ but reports by health care professionals and consumers is voluntary.³⁰⁶

Evolving from previous incarnations,³⁰⁷ 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.³⁰⁸ But the raw data files are published only in quarterly increments,³⁰⁹ 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.”)³¹⁰ FDA also instructs potential users to request individual case safety reports by submitting a FOIA request.³¹¹ 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.³¹²

³⁰⁰ FDA, Recalls, Market Withdrawals, & Safety Alerts, at <http://www.fda.gov/Safety/Recalls/default.htm>.

³⁰¹ FDA, Enforcement Report, at http://www.accessdata.fda.gov/scripts/ires/index.cfm#tabNav_advancedSearch.

³⁰² FDA, Drugs@FDA: FDA Approved Drug Products, at <http://www.accessdata.fda.gov/scripts/cder/drugsatfda/index.cfm>.

³⁰³ FDA, FDA Adverse Drug Event Reporting System (FAERS),

³⁰⁴ 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>).

³⁰⁵ 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).

³⁰⁶ FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS), at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

³⁰⁷ 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).

³⁰⁸ FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS), at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

³⁰⁹ FDA, FDA Adverse Event Report System (FAERS), Latest Quarterly Data Files, at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

³¹⁰ FDA, Questions and Answers on FDA's Adverse Event Reporting System (FAERS), at <http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/ucm082193.htm>.

³¹¹ *Id.*

³¹² FDA, MedWatch: The FDA Safety Information and Adverse Event Reporting Program, at

The device counterpart to FAERS is MAUDE (Manufacturer and User Facility Device Experience), a database of device adverse events.³¹³ The agency began requiring device reports in 1984.³¹⁴ 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.³¹⁵ The MAUDE database allows users to search for medical devices that may have malfunctioned or caused death or serious injury.³¹⁶ Users can search by a pull-down menu of product problems, by the class of product, or by manufacturer, model, or brand name.³¹⁷ Like FAERS, MAUDE is designed to help the FDA monitor emerging product safety problems, but unlike FAERS, the centralized search function makes the data accessible to lay users.

Like other agencies, the FDA is publishing FAERS, MAUDE, and other databases in more user-friendly formats on its Open FDA site.³¹⁸ 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.³¹⁹ 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 2004.³²⁰ Open FDA includes extensive data tools and downloadable raw data files, obviously directed at third-party users.

As with 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 of data, including “public, academic, and private entities” to “link and analyze safety data from multiple sources,” with an idea of uncovering emerging product safety risks.³²¹ The goal was to include at least 100 million patients in the dataset by 2012.³²² The FDA’s Sentinel Initiative 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.³²³ A pilot version (called “Mini-Sentinel”) is now available online,³²⁴

<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 JAMA 2765 (Jun. 2, 1993).

³¹³ FDA, MAUDE, at

<https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfmaude/search.cfm>.

³¹⁴ FDA, Medical Device Reporting, 49 Fed. Reg. 36,326 (Sep. 14, 1984).

³¹⁵ FDA, Mandatory Reporting Requirements: Manufacturers, Importers, and Device User Facilities, at

<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/PostmarketRequirements/ReportingAdverseEvents/ucm127891.htm>.

³¹⁶ *Id.*

³¹⁷ FDA, MAUDE – Manufacturer and User Facility Device Experience, at

<http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMAUDE/search.CFM>.

³¹⁸ Cortez, ACUS Report, *supra* note ___, at 44; FDA, OpenFDA, at <https://open.fda.gov>.

³¹⁹ Cortez, ACUS Report, *supra* note ___, at 44; FDA, OpenFDA, at <https://open.fda.gov>.

³²⁰ FDA, OpenFDA, Drugs – Adverse Events, at <https://open.fda.gov/drug/event/>.

³²¹ 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.).

³²² 21 U.S.C. § 355(k)(3)(B).

³²³ 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,

including a search tool for the Mini-Sentinel Distributed Database.³²⁵

Like other agency databases, FDA databases are not always complete and accurate. A 2011 study found widespread errors and incomplete reports filed in MedWatch, including more than 25% of reports using inaccurate product names.³²⁶

More than most agencies, the FDA includes prominent disclaimers of the accuracy and reliability of its data.³²⁷ For example, the MAUDE database includes, right below the search fields, several disclaimers, including the following:

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.³²⁸

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

First, there is no certainty that the reported event (adverse event or medication error) was actually due to the product. FDA does not require that a causal relationship between a product and event be proven, and reports do not always contain enough detail to properly evaluate an event. Further, FDA does not receive reports for every adverse event or medication error that occurs with a product. Many factors can influence whether or not an event will be reported, such as the time a product has been marketed and publicity about an event. Therefore, FAERS data cannot be used to calculate the incidence of an adverse event or medication error in the U.S. population.³²⁹

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

65 FOOD & DRUG L.J. 67 (2010) (considering whether participation in the Sentinel Initiative might be mandatory rather than voluntary).

³²⁴ FDA, Mini-Sentinel, at <http://mini-sentinel.org>.

³²⁵ FDA, Mini-Sentinel, Distributed Query Tool and Summary Tables, at http://mini-sentinel.org/data_activities/distributed_query_tool/default.aspx.

³²⁶ 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).

³²⁷ Cortez, ACUS Report, *supra* note ___, at 44.

³²⁸ 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).

³²⁹ FDA, FAERS, *supra* note ___.

unvalidated.”³³⁰

Still, 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.³³¹ Thus, an alliance between improving FDA and CMS data might help cure some defects in the FDA’s adverse event reporting systems.³³²

III. DISCLOSURE AND ITS ASPIRATIONS

These and other databases have become ubiquitous because disclosure appeals to so many of our intuitions about how government, markets, and regulation should work. So much so, in fact, that disclosure skeptics criticize 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.”³³³ Contemporary scholarship, of course, has focused on the many ways in which Internet technologies have facilitated communication between the government and the public.³³⁴ And after years of scholarly praise of the Internet’s role in facilitating transparency, accountability, and democracy, inevitably critiques have emerged.³³⁵ Still, disclosure frequently is offered as a tool that can achieve democratic, market, and regulatory ideals.

A. *Democratic Ideals*

A primary justification for disclosure policies is to enhance the accountability of government.³³⁶ And a value closely linked to accountability is transparency.³³⁷ In fact, government transparency is often assumed to be a precondition for accountability.³³⁸ Leading theorists—from John Milton, John Stuart Mill, James Madison, Oliver Wendell Holmes (father of the marketplace of ideas),³³⁹ and Louis Brandeis—have long drawn an explicit link between transparency, accountability, and democracy.³⁴⁰ Centuries of writing is filled with glittering

³³⁰ See, e.g., FDA, Open FDA, Drugs – Adverse Events, at <https://open.fda.gov/drug/event/>.

³³¹ See, e.g., Scott Gottlieb, *Opening Pandora’s Pillbox: Using Modern Information Tools to Improve Drug Safety*, 24 HEALTH AFF. 938 (2005).

³³² *Id.*

³³³ Ben-Shahar & Schneider, *supra* note ___, at 734.

³³⁴ 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); [OTHERS].

³³⁵ See, e.g., Matthew Hindman, *The Myth of Digital Democracy* (2003); [OTHERS].

³³⁶ Shkabatur, *supra* note ___, at 80; Adam M. Samaha, *Government Secrets, Constitutional Law, and Platforms for Judicial Intervention*, 53 UCLA L. REV. 909, 917 (2006); [OTHERS].

³³⁷ Open Government Memorandum, at 4685.; Shkabatur, *supra* note ___, at 83.

³³⁸ Open Government Memorandum, at 4685; Fenster, *supra* note ___, at 894-99; Shkabatur, *supra* note ___, at 83.

³³⁹

³⁴⁰ 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.

generalities about the virtues of transparency. Today, the modern open source movement, which owes much to these forebears, holds that information is a necessary precondition for truth and progress.³⁴¹

Government databases can also serve an expressive function by acting as a conduit for consumers to air their grievances against regulated firms.³⁴² Presidents Kennedy and Nixon both promulgated a Consumer Bill of Rights to pose the government as an intermediary or tribunal through whom consumer complaints could be given voice.³⁴³ 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."³⁴⁴ The CFPB's Consumer Complaint Database, as described above, successfully serves this role today.

B. *Market Ideals*

Another frequently-invoked rationale for agency disclosure, and in fact a rationale underpinning many mandatory disclosure regimes, is that disclosure can improve consumer decisionmaking, facilitate markets, and "protect the naïve from the sophisticated."³⁴⁵ 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 ultimately undermine the market.³⁴⁶ By requiring disclosure, the government can correct such information asymmetries and facilitate efficient markets. Schauer calls this "transparency as efficiency"—the idea that freely available information "is precisely what makes markets operate effectively."³⁴⁷

Agency databases might generate 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.³⁴⁸ The Bureau points to this phenomenon in the airline industry,

³⁴¹ 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.*

³⁴² 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).

³⁴³ *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.

³⁴⁴ Porter, *supra* note ___, at 76.

³⁴⁵ Omri Ben-Shahar & Carl E. Schneider, *The Failure of Mandated Disclosure*, 159 *U. PA. L. REV.* 647, 649 (2011).

³⁴⁶ 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).

³⁴⁷ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1350.

³⁴⁸ CFPB, *Disclosure of Certain Credit Card Complaint Data*, 76 *Fed. Reg.* 76,628, 76,630 n.9

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 also make use of government data.³⁴⁹ The Bureau concludes that after the data is made public, “The marketplace of ideas then does the rest.”³⁵⁰

Disclosure thus satisfies both our free-market intuitions³⁵¹ and the policymaker’s urge to do something. In the law that created the CFPB, Congress repeatedly asserts that the Bureau will publish information that helps consumers make more informed choices about financial products and services³⁵²—a refrain repeated frequently by the Bureau in its publications.³⁵³ Consumer advocates, in fact, have encouraged the Bureau’s databases by arguing that “disclosure is one of the best tools government agencies can use.”³⁵⁴

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.”³⁵⁵ Countless laws rely on this logic to justify mandatory disclosures.³⁵⁶ The CFPB’s complaint data, the CPSC’s product safety data, and countless other data sets follow this logic.

Finally, corporations and industry groups often “urge greater transparency as an alternative to allegedly more heavy-handed regulation.”³⁵⁷ Of course, many scholars embrace this view as well.³⁵⁸ As 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.”³⁵⁹ Regulation by disclosure thus appeals across political and ideological spectra.

C. *Regulatory Ideals*

A third aspiration of disclosure is to achieve regulatory ends—using publication to preempt undesired behavior or at least as a deterrent. Corporate and securities law, for example, relies heavily on disclosure of company financial reports and

(Dec. 8, 2011).

³⁴⁹ 76 Fed. Reg. at 76,631.

³⁵⁰ 76 Fed. Reg. at 76,631.

³⁵¹ Ben-Shahar & Schneider, *supra* note ___, at 681.

³⁵² 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”).

³⁵³ *See, e.g.*, 78 Fed. Reg. 21,225.

³⁵⁴ 78 Fed. Reg. at 21,220.

³⁵⁵ Ben-Shahar & Schneider, *supra* note ___, at 681.

³⁵⁶ Ben-Shahar & Schneider, *supra* note ___, at 681 (citing examples from the FTC and others).

³⁵⁷ 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).

³⁵⁸ *See, e.g.*, Crovitz, *supra* note ___; Thaler & Sunstein, *supra* note ___.

³⁵⁹ FUNG ET AL., *supra* note ___, at 5.

transactions, with the idea that corporations whose dealings are transparent and publicly accessible will think twice before acting in ways that damage investors.³⁶⁰ Similarly, requiring hospitals to publish mortality rates is really a device to encourage hospitals to reduce mortality rates.³⁶¹ Again, the same logic undergirds countless disclosure requirements. This is not a new idea—Jeremy Bentham observed in 1796 that “the more strictly we are watched, the better we behave.”³⁶²

Agencies also frequently use databases to publish compliance and enforcement data. Scholars have called for agencies to publish their enforcement records, which are available under FOIA, but must be requested.³⁶³ 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.”³⁶⁴ 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).³⁶⁵

Regulation by database can also help counter underenforcement by agencies, what Matthew Stephenson calls “agency slack.”³⁶⁶ Scholars note widespread underenforcement by a variety of agencies, in a variety of contexts.³⁶⁷ And even when agencies do pursue regulatory violations, they often fail to enforce them.³⁶⁸ Underenforcement may derive from several sources—insufficient agency resources, ideology, anti-regulatory pressures, political oversight, inertia, or agency self-interest.³⁶⁹ Regardless of the contributors, making compliance and

³⁶⁰ 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).

³⁶¹ 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.

³⁶² 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.

³⁶³ Vladeck, *supra* note ___, at 1830-31.

³⁶⁴ Vladeck, *supra* note ___, at 1830.

³⁶⁵ 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 ___.

³⁶⁶ Matthew C. Stephenson, *Public Regulation of Private Enforcement*, 91 *VA. L. REV.* 93, 110 (2005).

³⁶⁷ 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).

³⁶⁸ 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).

³⁶⁹ Stephenson, *supra* note ___, at 110.

enforcement broadly public might encourage 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 using data from the FDA's drug and device adverse events databases (FAERS and MAUDE) and providing commercial products based on the data.³⁷⁰ 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.³⁷¹ Moreover, there is always the lingering fear that shareholders, plaintiffs' lawyers, media, bloggers, or other enforcement agencies use the 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.³⁷² Technologists have envisioned the government serving as a "platform" for innovation by providing data that inspires outside parties to create innovative uses of it.³⁷³ 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 formulate innovative, perhaps enlightening uses of the data.³⁷⁴ Examples of third-party uses of governmental data are becoming more frequent.

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.³⁷⁵ The site claims that the government has awarded "[m]ore than \$220 million in prize money" since 2010.³⁷⁶ The stated goal is to seek "innovative solutions from the public, bringing the best ideas and talent together to solve mission-centric problems."³⁷⁷ Indeed, even the administration's original Open Government Directive offered contests and prizes that incentivized the public to "tinker" with the information released.³⁷⁸ Thus, some of the most provocative uses of agency data may just be emerging.

In short, scholars have documented why disclosure regimes appeal 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

³⁷⁰ 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)).

³⁷¹ Cortez, ACUS Report, *supra* note ___, at 66.

³⁷² Crowdsourcing, Definition, Merriam-Webster Online Dictionary, at <http://www.merriam-webster.com/dictionary/crowdsourcing>.

³⁷³ 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.

³⁷⁴ Shkabatur, *supra* note ___, at 109.

³⁷⁵ About Challenge.gov, at <http://www.challenge.gov/about/>.

³⁷⁶ *Id.*

³⁷⁷ *Id.*

³⁷⁸ 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.

traditional regulation.³⁷⁹ Indeed, calls for disclosure and transparency are often justified as furthering the pursuit of truth, knowledge, and societal progress.³⁸⁰

IV. DISCLOSURE AND ITS DISCONTENTS

Does disclosure live up to its many promises? In this Part, I evaluate the shortcomings of disclosure, and then consider ways to design databases for more optimal disclosure in Part V.

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, for example) to less tangible (reputational harms).³⁸¹ Although scholarship on these harms is not voluminous, the harms are relatively well documented.³⁸² [Include examples from ACUS Report pp. 9-12.] 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.”³⁸³

Questions about the value of so-called “naming and shaming” have crept into various disciplines.³⁸⁴ In the book *Is Shame Necessary?*, Jennifer Jacquet considers the virtues and flaws of naming and shaming, and how social media and other modern modes of communication might scale shame effectively to change the behavior of corporations or even governments.³⁸⁵ Indeed, organizations like Wikileaks have derived tremendous power and influence from disclosure—the kind of non-governmental power normally reserved for mainstream media.³⁸⁶ Less sanguine views argue that “shaming is the very antithesis of the law.”³⁸⁷

A second objection to the use of disclosure as a regulatory tool is that it may be ineffective. There is growing scholarly skepticism that openness necessarily leads to knowledge, or that more information necessarily produces better decisions.³⁸⁸ 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.”³⁸⁹ And as Ben-Shahar and Schneider emphasize in their magisterial article, *The Failure of Mandated Disclosure*, disclosure regimes have a clear allure to policymakers, but often fail completely in meeting their goals,

³⁷⁹ Ben-Shahar & Schneider, *supra* note ___, at 681-82

³⁸⁰ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1350.

³⁸¹ See, e.g., Ernest Gellhorn, *Adverse Publicity by Administrative Agencies*, 86 HARV. L. REV. 1380 (1973); Nathan Cortez, *Adverse Agencies by Administrative Agencies in the Internet Era*, 2011 BYU L. REV. 1371 (2011); Cortez, ACUS Report, *supra* note ____.

³⁸² 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; [JOHN BRAITHWAITE BOOK].

³⁸³ Vladeck, *supra* note ___, at 1793 (noting, however, that secrecy should give way to publication, for example, when necessary to avoid death or serious injury).

³⁸⁴ [CITE BRAITHWAITE].

³⁸⁵ JENNIFER JACQUET, *IS SHAME NECESSARY: NEW USES FOR AN OLD TOOL* (2015).

³⁸⁶ Shkabatur, *supra* note ___, at 135-36.

³⁸⁷ Eric Posner, *A Terrible Shame: Enforcing Moral Norms Without Law Is No Way to Create a Virtuous Society*, SLATE.COM (Apr. 9, 2015).

³⁸⁸ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1351.

³⁸⁹ Vladeck, *supra* note ___, at 1792.

and in fact can have unintended consequences that hinder.³⁹⁰ Although their conclusions are challenged,³⁹¹ the virtues of disclosure no longer remain uncontested.

A third objection to agency disclosure in general, and databases in particular, is that they are not always complete or accurate. To be their most useful, online databases that purport to present truthful, objective information must do both. To borrow a phrase from computer science and software engineering, “Garbage in, garbage out.” Flawed inputs will 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 inputs.³⁹² And as the litigation challenging inaccurate reports on SaferProducts.gov demonstrates, agencies sometimes post inaccurate data, and sometimes that data can linger online before it is corrected or retracted.³⁹³ The FEC’s campaign finance data is widely acknowledge to be incomplete.³⁹⁴ Even relatively non-controversial databases like USAspending.gov and Recovery.gov have been found to have widespread inaccuracies.³⁹⁵ Both the OMB and GAO have admitted that data on USAspending.gov has been inaccurate, untimely, or incomplete.³⁹⁶ An independent review by the Sunlight Foundation found that the site had “over 1.2 trillion dollars’ worth of misreported spending in 2009 alone.”³⁹⁷

A related criticism of agency databases is that they often present data without appropriate context.³⁹⁸ 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.³⁹⁹ 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 observe that agencies have significant discretion to decide what

³⁹⁰ Ben-Shahar & Schneider, *supra* note ____.

³⁹¹ Craswell, *supra* note ____.

³⁹² Timur Kuran & Cass R. Sunstein, *Availability Cascades and Risk Regulation*, 51 STAN. L. REV. 683 (1999).

³⁹³ 900 F. Supp. 2d 572.

³⁹⁴ See Part II.A, *supra*.

³⁹⁵ See, e.g., Shkabatur, *supra* note ____, at 103.

³⁹⁶ 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).

³⁹⁷ 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.

³⁹⁸ 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).

³⁹⁹ Shkabatur, *supra* note ____, at 104-05 (noting that these are two real-life examples in USAspending.gov).

data to disclose, and the scope of that data.⁴⁰⁰ 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.⁴⁰¹ Industry commenters also objected that the complaints lacked context, and that users might overlook the data's limitations, despite disclaimers by the Bureau.⁴⁰²

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 194,000 data sets,⁴⁰³ some of which include millions of records or data points. But more information is not always better information—although many disclosure policies assume otherwise.⁴⁰⁴ Federal agencies post so much information online today that some scholars wonder who is served by these data dumps.⁴⁰⁵ 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.”⁴⁰⁶ Genuinely useful information is often buried.⁴⁰⁷ As Paredes notes, sunlight can be the best disinfectant, “[b]ut sunlight can also be blinding.”⁴⁰⁸

So how does salient information stand out? Scholars have analyzed mandatory disclosures regimes, finding 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.”⁴⁰⁹ 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.”⁴¹⁰ As Ben-Shahar and Schneider observe, consumers “encounter too many disclosures to digest most of them.”⁴¹¹ 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.”⁴¹²

Another shortcoming of disclosure is the complexity of the data. Mandated disclosure regimes have become ubiquitous in federal and state law, ensconced in statutes, ordinances, agency regulations, and common law.⁴¹³ Sometimes, these

⁴⁰⁰ Shkabatur, *supra* note ___, at 117.

⁴⁰¹ See, e.g., 77 Fed. Reg. at 37,561.

⁴⁰² See, e.g., 77 Fed. Reg. at 37,562.

⁴⁰³ Data.gov, <http://www.data.gov> (last visited March 23, 2016).

⁴⁰⁴ Ben-Shahar & Schneider, *supra* note ___, at 650.

⁴⁰⁵ Shkabatur, *supra* note ___, at 118.

⁴⁰⁶ Vladeck, *supra* note ___, at 1832.

⁴⁰⁷ Ben-Shahar & Schneider, *supra* note ___, at 737.

⁴⁰⁸ Paredes, *supra* note ___, at 417.

⁴⁰⁹ 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.

⁴¹⁰ Ben-Shahar & Schneider, *supra* note ___, at 679.

⁴¹¹ Ben-Shahar & Schneider, *supra* note ___, at 705.

⁴¹² Shkabatur, *supra* note ___.

⁴¹³ Ben-Shahar & Schneider, *supra* note ___, at 650.

sources of law demand “marvelously elaborate disclosures” that are difficult if not impossible for the intended beneficiaries (usually consumers) to understand.⁴¹⁴ Consumers, of course, are not perfectly rational, but exercise “bounded rationality” due to various cognitive biases and distortions.⁴¹⁵ Thus, even in a world free from information asymmetries, our decisions are not perfectly rational. 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....”⁴¹⁶ 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.⁴¹⁷

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.”⁴¹⁸ 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,⁴¹⁹ which often goes unread.⁴²⁰ 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.”⁴²¹ 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.”⁴²² Any thoughtful disclosure regime must determine what information must be disclosed, by whom, to whom, the optimal format, and appropriate enforcement mechanisms.⁴²³ 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.⁴²⁴ For regimes with low rates of compliance,

⁴¹⁴ Ben-Shahar & Schneider, *supra* note __, at 650.

⁴¹⁵ 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).

⁴¹⁶ FUNG ET AL., *supra* note __, at 34.

⁴¹⁷ Craswell, *supra* note __, at 361.

⁴¹⁸ Ben-Shahar & Schneider, *supra* note __, at 665.

⁴¹⁹ The Jeanne Clery Disclosure of Campus Security Policy and Campus Crime Statistics Act, Pub. L. No. _____, *codified at* 20 U.S.C. § 1092(f).

⁴²⁰ Ben-Shahar & Schneider, *supra* note __, at 702-04; Sara Lipka, *Do Crime Statistics Keep Students Safe?*, CHRON. HIGHER EDUC., Jan. 30, 2009, at A1;

⁴²¹ 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.

⁴²² FUNG ET AL., *supra* note __, at 39.

⁴²³ FUNG ET AL., *supra* note __, at 39-46.

⁴²⁴ FUNG ET AL., *supra* note __, at 45 (examining civil and criminal penalties to enforce the disclosure mandates in campaign finance law and corporate law).

policymakers have often increased both penalties and enforcement efforts.⁴²⁵

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 each, at an estimated cost of over \$89 million per year.⁴²⁶ Another recent study found that U.S. physicians in just four common specialties spend \$15.4 billion annually reporting under various quality measurement programs.⁴²⁷ Thus, regulation by disclosure can be costly. [Add description from ACUS Report regarding how the CFPB created its database.]

The other costs are opportunity costs. Relying on disclosure as a means to pursue regulatory ends may mean bypassing other, better means for achieving those ends.⁴²⁸ Traditional command-and-control regulation long ago lost its luster among policymakers and academics.⁴²⁹ But modern replacements like “new governance,” despite their promise, 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.”⁴³⁰

Finally, there is some evidence that mandatory disclosure regimes can backfire. Consumers might let their guards down when presented with mandated disclosures that give transactions a “veneer of legality.”⁴³¹ 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.⁴³² Although there is much more research to be done, again, the virtues of disclosure are no longer uncontested.

V. DESIGNING FOR OPTIMAL DISCLOSURE

How can policymakers design databases that live up to the many aspirations of disclosure, while minimizing potential burdens? How can policymakers ensure the quality and reliability of agency databases?⁴³³ My first prescription is modesty. As Richard Craswell cautions, “people who expect disclosure laws to solve almost every problem—quickly, easily, and with very little cost—are doomed to have their expectations crushed.”⁴³⁴ The truth is that regulation by disclosure requires just as many difficult design and implementation choices as any other form of

⁴²⁵ FUNG ET AL., *supra* note __, at 45-46.

⁴²⁶ SEC, Form 8-K Disclosure of Certain Management Transactions, Securities Act Release No. 33-8090, 77 SEC Dock. 1072 (Apr. 12, 2002).

⁴²⁷ 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 north of \$2.6 trillion. [CITE].

⁴²⁸ Ben-Shahar & Schneider, *supra* note __, at 737-42.

⁴²⁹ [CITE NEW GOVERNANCE LITERATURE.]

⁴³⁰ Schauer, *Transparency in Three Dimensions*, *supra* note __, at 1348 (citing Ripken, *supra* note __).

⁴³¹ Ben-Shahar & Schneider, *supra* note __, at 740.

⁴³² Ben-Shahar & Schneider, *supra* note __, at 739 (citing such evidence related to conflict-of-interest disclosures).

⁴³³ 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.

⁴³⁴ Craswell, *supra* note __, at 379.

regulation.⁴³⁵ The early successes of the CFPB’s Consumer Complaint Database and the CPSC’s SaferProducts.gov site owe to several canny decisions by Congress and the agencies, as detailed above and incorporated below.

The most successful disclosure regimes carefully consider data inputs (how data will be collected and from whom), outputs (how the data will be published and presented), and procedures to safeguard the quality and reliability of the data. I thus offer recommendations in three forms—addressing data inputs, outputs, and procedural safeguards (including both pre- and post-publication procedures). Together, these recommendations propose a decidedly modern role for the government.

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, I try to highlight emerging best practices from the databases I have evaluated, in the hopes that these discussions will be useful to policymakers going forward. Again, these recommendations build on (and in some instances exceed) those I helped develop by ACUS.

A. *Inputs*

Policymakers should think carefully, and ideally in advance, about inputs—who will provide data to populate data sets, how, and the extent to which the agency itself will try to verify, validate, or authenticate the information.

First and foremost, agencies that purport to publish accurate and objective information should adopt procedures to these ends. As emphasized in Part III, “transparency” is frequently invoked to support disclosure, but it is important to remember that ‘transparency,’ as defined in dictionaries, allows objects to be “seen without distortion.”⁴³⁶ Thus, government databases should endeavor to achieve genuine transparency, rather than translucency.⁴³⁷ Indeed, some scholars note that inevitably, disclosure regimes created through the political process are forged by conflict and compromise, and thus generate only partial transparency, rather than full disclosure.⁴³⁸ But scholars also find that more successful disclosure regimes tend to increase the accuracy and quality of information they publish over time.⁴³⁹ Databases thus do not have to be static; they can be dynamic. Although initial design choices are important, agencies should not hesitate to tinker if flaws in the data become apparent.

Of course, a primary consideration for agencies is to identify appropriate data sources. The irony of the Internet is that it 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 the reliability of such data⁴⁴⁰ Thus,

⁴³⁵ Craswell, *supra* note ___, at 379.

⁴³⁶ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1343 (referring to the *Oxford English Dictionary*’s definition of “transparency”).

⁴³⁷ Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1345.

⁴³⁸ FUNG ET AL., *supra* note ___, at xii.

⁴³⁹ FUNG ET AL., *supra* note ___, at 109.

⁴⁴⁰ *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.”).

government agencies can play an important role by ensuring that published data is credible.⁴⁴¹ In fact, because publication by a government agency itself signals credibility, it is incumbent on agencies to ensure the accuracy of data that purports to be objective, or adequately explain the source and any limitations on data that does not.⁴⁴² Databases should be labeled and characterized accurately, much as we expect from industry labeling.⁴⁴³ This is particularly so because government agencies are perhaps the most trusted source of information, and information in agency databases often carry the imprimatur of the federal government.⁴⁴⁴ Thus, the sources of data should be clearly stated, including whether the data are contested, and any steps the agency takes (or more importantly does not) to resolve such contests.⁴⁴⁵

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.⁴⁴⁶ The CFPB, for example, does not verify that consumer complaints are “accurate” (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 identified, and gives the company ample opportunity to identify false or fraudulent complaints.⁴⁴⁷ Likewise, when Congress authorized SaferProducts.gov, it required the CPSC to consider objections that the information is “materially inaccurate.”⁴⁴⁸ Both the CFPB and CPSC allow the subjects of complaints to post responses, which are published alongside complaints.⁴⁴⁹

Reliable databases depend not just on publishing accurate data (if that is what they purport to do), but also on publishing relatively complete and representative data. Industry commenters objected repeatedly to the CFPB that its database of self-selected consumer complaints would necessarily be incomplete, non-random, and thus non-representative of company performance and consumer experiences.⁴⁵⁰ 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.”⁴⁵¹

A related concern is that data selection or filtering criteria might publish data that are technically accurate but misleading as a whole. Transparency initiatives often fail when “transparency is either not sufficiently mandatory or not applicable to categories of information that meaningfully contribute to public accountability.”⁴⁵² 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.”⁴⁵³ The data

⁴⁴¹ Heerwig & Shaw, *supra* note __ 1476.

⁴⁴² Cortez, ACUS Report, *supra* note __, at __.

⁴⁴³ [Cite, e.g., FDA labeling authority.]

⁴⁴⁴ Cortez, ACUS Report, *supra* note __, at 94.

⁴⁴⁵ Cortez, ACUS Report, *supra* note __, at 96.

⁴⁴⁶ Madison, *supra* note __, at 1607-08.

⁴⁴⁷ [CITE].

⁴⁴⁸ 15 U.S.C. § 2055(c)(4).

⁴⁴⁹ Cortez, ACUS Report, *supra* note __, at 20-21.

⁴⁵⁰ See, e.g., 77 Fed. Reg. at 37,561.

⁴⁵¹ *Id.*

⁴⁵² Shkabatur, *supra* note __, at 105-06.

⁴⁵³ Adam Samaha, Government Secrets, Constitutional Law, and Platforms for Judicial

chosen for publication may not paint a complete picture. Thus, Shkabatur argues, the answer for incomplete transparency is even more transparency, and more enforcement of those requirements. 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.⁴⁵⁴

Agencies must also think carefully about the format in which they will publish data. The trend, as noted above, is to publish data sets both in 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 includes normative judgment calls that might draw into question how objective or neutral the presentation is. But the latter is a relatively recent phenomenon. Historically, agencies have been reluctant to publish information in open or “raw” data format (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.”⁴⁵⁵ Yet, as far back as 2004, the OMB encouraged agencies to “provide all data in an open, industry standard format permitting users to aggregate, disaggregate, or otherwise manipulate and analyze the data to meet their needs.”⁴⁵⁶ 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.⁴⁵⁷ Thus, there is clear trend toward publishing in raw, open data formats.

At the same time, some scholars argue that publishing data in raw, “naked” formats can itself serve as a barrier for non-programmers and others who are not able to understand or use such data.⁴⁵⁸ Raw government datasets might require, ironically, technically sophisticated intermediaries to decipher.⁴⁵⁹ 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.

Nevertheless, some argue that the government should focus its energies less on presenting packaged information on its web sites and more on publishing

Intervention, 53 UCLA L. REV. 909, 919 (2006); Shkabatur, *supra* note ___, at 106.

⁴⁵⁴ See, e.g. Vladeck, *supra* note ___, at 1828-29.

⁴⁵⁵ 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).

⁴⁵⁶ 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>.

⁴⁵⁷ 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).

⁴⁵⁸ See, e.g., Shkabatur, *supra* note ___, at 112.

⁴⁵⁹ *Id.* at 112.

“reusable data.”⁴⁶⁰ 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.⁴⁶¹ Robinson and colleagues argue that nongovernmental users will be better able to experiment with how best to present the data, whether it be with advanced search functionalities, automated content analysis, indexing among multiple sources, and various data visualization tools.⁴⁶² They call this new role for agencies an “invisible hand,” enabling a “marketplace of engineering ideas.”⁴⁶³ Some will also value being able to access “genuine” data that is not mediated, framed, or translated by an intermediary (including the government).⁴⁶⁴

We might be experiencing a major shift in the government’s informational role from controller to facilitator. Modern agencies may be best suited to facilitating rather than controlling information—that is, gathering and publishing data, and ensuring its quality, enabling the private and non-profit sectors to maximize its use.⁴⁶⁵ Data users might also become contributors, as in the case of consumers who report food poisoning from restaurants, who are able to supplement relatively infrequent government inspections.⁴⁶⁶ Thus, like the CFPB’s aspirations, 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.⁴⁶⁷ Such disclosures are mandated by government but are also integrated by disclosers because the information is useful to consumers or users.⁴⁶⁸ Another factor that improves success is whether the information varies between disclosers, such that disclosers have an incentive to race to the top.⁴⁶⁹ 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 do have an incentive to reduce the tar and nicotine content of their products, which varies from brand to brand.⁴⁷⁰ Thus, data can serve as an important differentiator between competitors, and they might spend their own resources publicizing their data.⁴⁷¹ 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

⁴⁶⁰ David Robinson, Harlan Yu, William P. Zeller, & Edward W. Felten, *Government Data and the Invisible Hand*, 11 YALE J.L. & TECH. 160, 160 (2009).

⁴⁶¹ Robinson *et al.*, *supra* note __, at 161.

⁴⁶² Robinson *et al.*, *supra* note __, at 161, 169.

⁴⁶³ Robinson *et al.*, *supra* note __, at 161.

⁴⁶⁴ 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.

⁴⁶⁵ See, e.g., FUNG ET AL., *supra* note __, at 166.

⁴⁶⁶ FUNG ET AL., *supra* note __, at 166.

⁴⁶⁷ Craswell, *supra* note __, at 369.

⁴⁶⁸ Craswell, *supra* note __, at 369-70.

⁴⁶⁹ Craswell, *supra* note __, at 371.

⁴⁷⁰ Craswell, *supra* note __, at 371.

⁴⁷¹ Craswell, *supra* note __, at 370-71.

health records. Through various pieces of legislation,⁴⁷² 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.⁴⁷³ 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.⁴⁷⁴ The goal is to put open, machine-readable data in the hands of programmers, entrepreneurs, journalists, providers, scientists, consumers, and other policymakers who might, in turn, help improve health care in the United States.⁴⁷⁵ 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.⁴⁷⁶

Thinking carefully about design choices and database inputs can help achieve some of the more realistic aspirations for disclosure.

B. *Outputs*

Policymakers should also consider the “outputs” of agency databases—how the data will be published, presented, and used. Should agencies rely on massive raw data dumps targeted at more sophisticated users? Or should they package and distill the data for lay users?

Disclosure enthusiasts often assume that more is better. But recently, scholars questioning the value of disclosure have begun to recognize that it is more important that information be accessible and *usable*, rather than simply available.⁴⁷⁷ Perhaps the relevant question, then, is not what policymakers think consumers *need* to know, but what consumers *want* to know.⁴⁷⁸ Disclosure policies that consider what information users want and need, and what they can comprehend, tend to be more successful.⁴⁷⁹

And perhaps consumers do not need more data, but more advice.⁴⁸⁰ The opposite of making data “bigger,” of course, is making data “smaller” and easier to process—usually by simplifying, tailoring, and targeting the information.⁴⁸¹ Thus, rating systems and other information available at the point of purchase (or

⁴⁷² 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.).

⁴⁷³ 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).

⁴⁷⁴ HealthData.gov, Datasets, <http://www.healthdata.gov/dataset?f%5B0%5D=type%3Adataset> (sorted by “Publisher”).

⁴⁷⁵ HealthData.gov, About, <http://www.healthdata.gov/content/about>.

⁴⁷⁶ Madison, *supra* note ___, at 1620-21.

⁴⁷⁷ See, e.g., Schauer, *Transparency in Three Dimensions*, *supra* note ___, at 1344; Fenster, *supra* note ___, at 942.

⁴⁷⁸ Ben-Shahar & Schneider, *supra* note ___, at 746.

⁴⁷⁹ FUNG ET AL., *supra* note ___, at 11.

⁴⁸⁰ Ben-Shahar & Schneider, *supra* note ___, at 746.

⁴⁸¹ 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).

ideally slightly before) may be more useful for consumers.⁴⁸² Mere publication on a government web site may not be enough, 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.⁴⁸³ 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.⁴⁸⁴ Restaurant hygiene grades, for example, are thus more embedded in the decision of where to eat than complex and voluminous patient safety disclosures.⁴⁸⁵

Although “big data” is all the rage for agencies, there are emerging appeals for simplification. Thus, agencies are thinking more carefully today about ensuring the “utility” of data, perhaps owing in part to the Information Quality Act (IQA).⁴⁸⁶ But our notions of “utility” are evolving, leading us to question whether more data is always better. 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.⁴⁸⁷ Still, as Ben-Shahar and Schneider note, even with nutrition labeling, “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.⁴⁸⁸ But how much should this deter agency disclosure efforts?

Ratings systems that communicate data that has been simplified and translated have enjoyed moderate success.⁴⁸⁹ For example, there is evidence that simple letter grades for restaurant sanitation and hygiene (from “A” to “C”) have led to cleaner restaurants in Los Angeles County.⁴⁹⁰ Researchers found that 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.⁴⁹¹ More tellingly, studies show that prominent disclosure of these grades encouraged restaurants to improve their sanitation practices, which correlated with a significant local drop in hospitalizations caused by food-borne illnesses.⁴⁹² Thus,

⁴⁸² FUNG ET AL., *supra* note __, at 57.

⁴⁸³ FUNG ET AL., *supra* note __, at 65-74.

⁴⁸⁴ FUNG ET AL., *supra* note __, at 57.

⁴⁸⁵ FUNG ET AL., *supra* note __, at 65.

⁴⁸⁶ 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.

⁴⁸⁷ Ben-Shahar & Schneider, *supra* note __, at 675.

⁴⁸⁸ 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)).

⁴⁸⁹ Archong Fung *et al.*, Full Disclosure: The Perils and Promise of Transparency (2007).

⁴⁹⁰ 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).

⁴⁹¹ FUNG ET AL., *supra* note __, at 50 (citing studies).

⁴⁹² 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).

simple, comprehensible, and easily accessible ratings based on larger data points not only allowed consumers to vote with their wallets, but also encouraged restaurants to compete based on cleanliness—undoubtedly the underlying motivation.⁴⁹³

For disclosure policies to truly succeed, they must affect not only the decisionmaking of consumers and regulatory beneficiaries, but also the decisionmaking of the discloser—the regulated party.⁴⁹⁴ Thus, effective disclosure systems become “doubly embedded.”⁴⁹⁵ The way disclosure policies affect discloser behavior is intuitive—by affecting their profits, market share, and/or reputation.⁴⁹⁶ 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: a massive raw data dump intended for sophisticated intermediaries; or a highly distilled presentation intended for lay users. The correct choice depends very much on the database 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 or five-star crash safety ratings, which themselves are based on complex engineering standards and test results.⁴⁹⁷ Ratings and grades also combat the problem of overdisclosure. Scholars scrutinizing the effectiveness of mandatory disclosure regimes sometimes observe that parties can “overdisclose” information to try to “overwhelm and distract” the intended audience.⁴⁹⁸ Ratings help solve for that.

On the other hand, other efforts fail to make the complex understandable, as evidenced by the vague five-color scheme for communicating the threat of a terrorist attack.⁴⁹⁹ 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.⁵⁰⁰ Thus, not all data is amenable to such translation.

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.” To maximize accessibility, the data should be available in multiple formats, as many agencies now recognize, and as Data.gov demonstrates.

But use by third party information intermediaries is compelling for several reasons. Publicly-minded watchdogs like Pro Publica, the Sunlight Foundation, and the Center for Effective Government (formerly OMB Watch) can serve a

⁴⁹³ FUNG ET AL., *supra* note ___, at 50-51.

⁴⁹⁴ FUNG ET AL., *supra* note ___, at 65-74.

⁴⁹⁵ FUNG ET AL., *supra* note ___, at 65.

⁴⁹⁶ FUNG ET AL., *supra* note ___, at 66.

⁴⁹⁷ FUNG ET AL., *supra* note ___, at 59-61.

⁴⁹⁸ 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).

⁴⁹⁹ FUNG ET AL., *supra* note ___, at 61.

⁵⁰⁰ FUNG ET AL., *supra* note ___, at 61.

translational role, sifting large amounts of data into more understandable bits.⁵⁰¹ 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 smart intermediaries. For example, various consumer groups have tried to translate toxic release data into more user-friendly web sites.⁵⁰²

However, translating voluminous and 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 relevant industry). But the number of organizations that can fit comfortably into this 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.⁵⁰³ Moreover, some doubt that these organizations derive their value from information supplied by the government, rather than from their own surveys and information-collecting activities.⁵⁰⁴ Still, the fact that individuals like Joshua Tauberer (who created Govtrack.us in his spare time)⁵⁰⁵ and Carl Malamud (who painstakingly made SEC data available online),⁵⁰⁶ demonstrate that the barriers are far from insurmountable. Moreover, these extraordinary individual efforts inspired the government to publish the data in open formats.⁵⁰⁷

As agencies have discovered, data is becoming a more collaborative endeavor. There is optimism that once raw data is published, the private, public, and nonprofit sectors will compete to make the data accessible and useful to their respective constituents.⁵⁰⁸ 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.⁵⁰⁹

The new generation of disclosure will be more collaborative, in the sense that various sectors will both contribute to and use the data.⁵¹⁰ Indeed, there seems to be wide agreement that the government should not have a monopoly on generating data,⁵¹¹ but can play an important centripetal role in centralizing it and helping assure its quality.

C. Procedural Safeguards

⁵⁰¹ See, e.g., Shkabatur, *supra* note ___, at 118.

⁵⁰² 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>;

⁵⁰³ Robinson *et al.*, *supra* note ___, at 173.

⁵⁰⁴ Ben-Shahar & Schneider, *supra* note ___, at 731-32.

⁵⁰⁵ Robinson *et al.*, *supra* note ___, at 165-66, 171.

⁵⁰⁶ Robinson *et al.*, *supra* note ___, at 166, 171.

⁵⁰⁷ Robinson *et al.*, *supra* note ___, at 171.

⁵⁰⁸ FUNG ET AL., *supra* note ___, at 153.

⁵⁰⁹ FUNG ET AL., *supra* note ___, at 158-60.

⁵¹⁰ FUNG ET AL., *supra* note ___, at 153.

⁵¹¹ See, e.g., Esty, *supra* note ___, at 199-200.

In addition to data inputs and outputs, it is important for policymakers to consider procedural safeguards to help ensure the quality and reliability of the data published. This Part, then, recommends various pre- and post-publication procedures for agency databases. The appropriate safeguards, of course, depend on the nature of the database.

Data sets that purport to publish accurate, objective information should include, ideally, both pre- and post-publication procedures that allow parties to request that the 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.⁵¹² For example, by statute the CPSC must allow companies whose products are reported to SaferProducts.gov the opportunity to comment on Reports of Harm.⁵¹³ The CPSC must consider objections that the information is “materially inaccurate,”⁵¹⁴ and the Commission publishes these procedures in the C.F.R.⁵¹⁵ Likewise, the CFPB authenticates that complaints are coming from actual customers of the companies cited. Bureau procedures allow companies to use an online Company Portal to verify a commercial relationship with the customer, and post the company’s response.⁵¹⁶ The Bureau also makes clear that each complaint, before being published in the database, must meet several publication criteria.⁵¹⁷

Nevertheless, errors in published data sets are probably inevitable, no matter how robust the pre-publication procedures may be. Thus, policymakers should also consider post-publication procedures that help ensure the quality and reliability of the 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.⁵¹⁸

Here, the controversial Information or Data Quality Act might be of help. The Act required the OMB to publish government-wide guidelines for “ensuring and maximizing the quality, objectivity, utility, and integrity of information ... disseminated by the government.”⁵¹⁹ The Act applies broadly to “Federal agency dissemination of public information, regardless of the form or format.”⁵²⁰ 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.”⁵²¹

⁵¹² Cortez, ACUS Report, *supra* note ___, at 94-95.

⁵¹³ 15 U.S.C. § 2055(c)(2).

⁵¹⁴ 15 U.S.C. § 2055(c)(4).

⁵¹⁵ 16 C.F.R. § 1102.26.

⁵¹⁶ 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*.

⁵¹⁷ Cortez, ACUS Report, *supra* note ___, at 63-64.

⁵¹⁸ Gellhorn, *supra* note ___; Cortez, *supra* note ___; Cortez, ACUS Report, *supra* note ___ (Appendix C: Table of Federal Cases (1974-2014)).

⁵¹⁹ 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.

⁵²⁰ 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.

⁵²¹ IQA, *supra* note ___.

Per the OMB's guidelines, dozens of federal agencies have published their own such guidelines and post-publication procedures for correcting or retracting information.⁵²² 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."⁵²³ Second, they exclude "adjudicative processes."⁵²⁴ 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."⁵²⁵ 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.

What, however, should agencies do about data sets that do not purport to be accurate and objective? Agency databases are populated with a variety of data from a variety of sources, including consumers, regulated parties, or the agency itself—and each might require different quality controls and presentations.⁵²⁶ Agencies must be clear about their sources and any limitations on the data, including whether the data is contested. The FDA's adverse event reporting databases, for example, are populated by reports that a product may have been "associated" with an adverse event, without any claim as to causation.⁵²⁷ 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."⁵²⁸ Just like the FDA does not verify that a product caused a specific adverse event, the CFPB does not verify that a company even engaged in the conduct alleged in the complaint. Doctors subject to "report cards" also lament that death and complication rates are presented without being normalized for riskier patient populations.⁵²⁹ Should such "data" even be published?

For better or worse, routinely it is. The solution, perhaps, is for agencies to represent the data accurately, which often means explaining the sources of the data and any important statements about the context and any limitations of the data. Several agencies already try 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,

⁵²² See Office of Management and Budget, Agency Information Quality Guidelines, https://www.whitehouse.gov/omb/inforeg_agency_info_quality_links/.

⁵²³ 67 Fed. Reg. at 377.

⁵²⁴ *Id.*

⁵²⁵ 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>.

⁵²⁶ Cortez, ACUS Report, *supra* note ___, at 96.

⁵²⁷ See Part II.F, *supra*.

⁵²⁸ Cortez, ACUS Report, *supra* note ___, at 69-70; Porter, *The Complaint Conundrum*, *supra* note ___, at 78.

⁵²⁹ Sandeep Jauhar, *Giving Doctors Grades*, N.Y. TIMES (Jul. 22, 2015), <http://www.nytimes.com/2015/07/22/opinion/giving-doctors-grades.html>.

unverified, or biased data.”⁵³⁰ The FDA’s adverse drug event database also notes that “there is no certainty that the reported event ... was actually due to the product.”⁵³¹ 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.”⁵³² 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.”⁵³³ Although one court called the CPSC’s language “boilerplate” that “would not interest an ordinary consumer,”⁵³⁴ providing appropriate context for data and disclosing its limitations is relatively easy and helps answer several criticisms of disclosure noted in Part IV.

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.⁵³⁵

First, the OMB already performs a centripetal and centrifugal function. In addition to its Information Quality Act guidelines, the Paperwork Reduction Act requires the OMB to preapprove significant information gathering efforts by agencies.⁵³⁶ 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.⁵³⁷ 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 be reluctant to agree to 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,⁵³⁸ and the FDA’s many ombudsmen have fielded complaints under the Information Quality Act.⁵³⁹

A third option is review by Inspectors General. Inspectors General (IGs) are independent officers, directed by law to detect and prevent fraud, waste, and abuse in federal agencies.⁵⁴⁰ They also maintain, by design, crucial independence from agency heads, and thus can serve as an independent arbiter.⁵⁴¹ They also function as an avenue for industry complaints.⁵⁴² 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

⁵³⁰ FDA, MAUDE, *supra* note ____.

⁵³¹ FDA, FAERS, *supra* note ____.

⁵³² CFPB, Consumer Complaint Database, *supra* note ____.

⁵³³ 15 U.S.C. § 2055(b)(5).

⁵³⁴ *Company Doe v. Tenenbaum*, *supra* note ___, at 598.

⁵³⁵ Cortez, ACUS Report, *supra* note ___, at 100-02.

⁵³⁶ 44 U.S.C. § 3507(d).

⁵³⁷ *Id.* at 3508.

⁵³⁸ CFPB, Ombudsman’s Office, *Annual Report to the Director* (Nov. 17, 2014).

⁵³⁹ FDA, FDA’s Office of the Ombudsman: Dispute Resolution and Problem Solving, at 2.

⁵⁴⁰ Inspector General Act of 1978, Pub. L. No. 95-452, 92 Stat. 1101 (1978).

⁵⁴¹ 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).

⁵⁴² Cortez, ACUS Report, *supra* note ___, at 101.

completeness of the public complaint database.”⁵⁴³

Finally, agency Chief Information Officers might play an important role in not only designing database inputs and outputs, but also in participating in pre- and post-publication procedures described above.

Additional procedures thus hold great allure, though they are not without cost. Ironically, 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.”⁵⁴⁴ Although each separate requirement has its own logic, together they can limit how agencies present data, and generally favor standardization above experimentation.⁵⁴⁵ 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.”⁵⁴⁶ Again, nongovernmental intermediaries have proven useful in rendering government data more accessible and useful.⁵⁴⁷

Regardless of the pre- and post-publication procedures offered, they should be designed to be hospitable to consumers and regulatory beneficiaries—not just industry and regulated parties. Indeed, Fung, Graham, and Weil found that the most successful disclosure regimes “featured strong groups representing information users, offered benefits to at least some information disclosers, and provided comprehensible content.”⁵⁴⁸

CONCLUSION

Agency databases have proliferated on the belief that democracy, markets, and regulation 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 expectations, we also must recognize that sunlight can also blind and burn. It is in this spirit that I call for policymakers to embrace the government’s role as a data steward, a sentinel of quality that helps maximize the quality of data inputs and outputs. The more reliable government data are, the more enduring the innovations based on that data should be.

⁵⁴³ 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”).

⁵⁴⁴ Robinson *et al.*, *supra* note __, at 162.

⁵⁴⁵ Robinson *et al.*, *supra* note __, at 163-65.

⁵⁴⁶ Robinson *et al.*, *supra* note __, at 165.

⁵⁴⁷ Robinson *et al.*, *supra* note __, at 165.

⁵⁴⁸ FUNG ET AL., *supra* note __, at xiv.