
UCLA

JOURNAL OF LAW & TECHNOLOGY

Special Issue: Platforms and the Press

OPEN AND SHUT? THE PROMISE – AND PROBLEMS – OF GOVERNMENT OPEN DATA PORTALS IN MEETING COMMUNITY INFORMATION NEEDS

FRANK D. LOMONTE,¹ BRITTANY SUSZAN² AND PRIYA DAMES³

ABSTRACT

The open data revolution—a movement to liberate government-held information by publishing it online—holds enormous promise as a “force multiplier” for cash-strapped news organizations. Rather than consuming the resources of journalists and lawyers in fighting for access to government records, opening data voluntarily enables news organizations to devote their resources to adding value to government information through analysis and contextualization. But the reality of open data has yet to fully match its promise. This paper examines how and why, and recommends protocols to guide government agencies in selecting the highest-value datasets to publish.

Access to government data and documents is the fuel that powers investigative reporting. But the freedom-of-information (FOI) laws that entitle the public and press to government-held information are notoriously frustrating to use; compliance is often incomplete, costly and slow. Voluntarily opening high-value datasets to public inspection

¹ Senior Academic Fellow, Center for Governmental Responsibility, University of Florida Levin College of Law. Legal Counsel, Cable News Network, Inc. B.A., 1992, Political Science, Georgia State University; J.D., 2000, University of Georgia School of Law.

² Research Associate, Joseph L. Brechner Center for Freedom of Information, University of Florida. B.A., 2009, Business Management, Washington College.

³ Research Associate, Joseph L. Brechner Center for Freedom of Information. B.A., 2022, Political Science, University of Florida. M.S., 2024 (anticipated), Data Science, North Carolina A&T University.

can complement FOI laws in ways useful both to agencies (sparing them from repeat requests for the same information) and for requesters (sparing them from the adversarial process of suing for access).

The first generation of government open data portals, however, has failed to garner widespread public engagement. Researchers have suggested that part of the problem is agencies' failure to prioritize publishing the data that users actually need. To illustrate the shortcomings of municipal open data sites, the authors chose 30 cities of varying sizes and checked their websites for one of the highest-priority datasets in contemporary America: Instances in which police officers use force. Predictably, the review found that big cities – with the capacity to hire well-qualified information officers – were likely to publish the data, while small towns invariably did not.

Investing in open data is an investment in rebuilding frayed trust with a skeptical public. Reliable government data can be an asset in combating dis- and mis-information. But making this investment will require changing both government spending priorities and government custodians' widespread cultural predisposition toward secrecy.

The authors recommend that, for open data portals to realize their civic potential, government agencies should prioritize the data they choose to publish by considering three priorities: Urgency, actionability, and verifiability. Although there is considerable controversy over whether news organizations should accept direct government subsidies, it would be uncontroversial for government agencies to support quality journalism indirectly, by lowering the barriers to obtaining useful information.

TABLE OF CONTENTS

INTRODUCTION	96
I. THE STRUGGLES (AND POTENTIAL) OF LOCAL JOURNALISM	98
II. THE STATE OF OPEN DATA.....	102
A. Access Law 101.....	102
B. The Question to Liberate Data	106
C. The Use Case: Open Data as a Journalistic “Force Multiplier”	109
D. Police Use-of-Force Data: A Case Study.....	113
III. ARRESTED DEVELOPMENT: WHY OPEN DATA HAS UNDERACHIEVED	115
A. Prioritization and Political Will.....	117
B. Privacy Concerns.....	119
C. “Data Deserts”: Who’s Counting	123
D. Lack of Audience Engagement.....	125
IV. RECOMMENDATIONS.....	127
V. CONCLUSION	134

INTRODUCTION

Data is the fuel that powers journalistic coverage of public affairs. Data helps answer questions such as: Is crime getting better, worse, or staying the same? Are taxes being assessed and collected equitably, and if not, who bears the brunt of the inequity? Are standardized test scores increasing or declining, and how do those outcomes correspond with school demographics or funding levels? Without data, journalism can become a collection of potentially meaningless or even misleading anecdotes. But data has historically been challenging to assemble. Requesting records from government agencies is slow, costly, and often devolves into adversarial haggling.⁴

Enter the open data portal. Modern government agencies no longer wait to be asked for data about crime, taxes, or schools. Instead, agencies affirmatively publish data on a publicly owned and maintained platform.⁵ The promise of open data is immense: It has the capability to provide fast, conflict-free access to information that can empower the public to participate effectively in civic life and to discharge its watchdog oversight role.⁶ That, in any event, is the promise.

But the reality can fall disappointingly short. Government agencies do not reliably prioritize publishing the highest-value data sets that would meet the public's information needs. Even when high-value data is made available, it is typically published in summary form, without the underlying source documentation that would enable a

⁴ See David Cuillier, *Forecasting Freedom of Information*, KNIGHT FOUND. 7 (March 2017), <https://kf-site-production.s3.amazonaws.com/publications/pdfs/000/000/232/original/FOI-final-unlink.pdf> (reporting results of nationwide survey of journalists, lawyers and nonprofit open-government advocates, of whom 73.9 percent said that delayed fulfillment of requests for public records was “very” or “extremely” problematic, while 46.2 percent said that excessive redaction of records was “very” or “extremely” commonplace).

⁵ See Beth Simone Noveck, *Rights-Based and Tech-Driven: Open Data, Freedom of Information, and the Future of Government Transparency*, 19 YALE HUM. RTS. & DEV. L.J. 1, 4–5 (2017) (explaining that philosophy behind open data movement shifts relationship between government and user “from the adversarial to the collaborative” and, theoretically, advances participatory democracy).

⁶ See Micah Altman et al., *Towards a Modern Approach to Privacy-Aware Government Data Releases*, 30 BERKELEY TECH. L.J. 1967, 1970 (2015) (“Making government data more widely available promises to enhance organizational transparency, improve government functions, encourage civic engagement, support the evaluation of government decisions, and ensure accountability for public institutions.”).

skeptical consumer to verify it.⁷ And with little consistency among government agencies when reporting and recording data, making meaningful apples-to-apples comparisons among jurisdictions can be challenging.⁸

This paper seeks to synthesize the growing body of research about the value of affirmatively opening government data for public inspection by way of online portals and discusses where the first generation of open data initiatives has fallen short of its potential. Part I looks at the fraying of America's information safety net, as cash-strapped local news organizations downsize and close, and what that trend means for public access to trustworthy data. It discusses the growing adoption of computer-assisted reporting, which can amplify journalists' ability to analyze government data—if the data is made available in a usable format. Part II examines the current state of open data, with a focus on municipal portals, because municipalities have often been the innovators while agencies at other levels of government have lagged. Part II starts by examining the philosophy behind the open data movement and how it has evolved, and looks at the legal and regulatory framework that guides how data is gathered and shared. It incorporates findings from a review of municipal sites, concluding that one of the most obviously valuable data sets—details about police use of force, including how complaints of excessive force are resolved—is only spottily available, with discussion of the implications of that shortcoming. Part III turns to evaluating where municipal open data portals have struggled to realize their potential, and why government data-miners predictably may hesitate to share sensitive data, even when that data is of paramount civic value. Part IV draws on prior open data scholarship and advocacy by nongovernmental organizations to derive some best practice recommendations for those who

⁷ See Sara Kizer, *Visualizing a Way Forward: Lessons Learned and Successful Strategies From Covid-19 Data Dashboard Implementation*, 61 JURIMETRICS J., Summer 2021, at 469, 479 (commenting on unreliability of universities' COVID data dashboards, "many of these dashboards lack clear information about how metrics are calculated and how often the data are refreshed"); Ryan Little, *Monsters: Fancy Government Charts With No Raw Data*, MOTHER JONES (Dec. 27, 2021), <https://www.motherjones.com/politics/2021/12/monsters-fancy-government-charts-with-no-raw-data/> (commenting on journalists' frustration when government databases do not provide operability to download and analyze underlying raw data, making it impossible to "check their math")

⁸ See Carla Bonina & Ben Eaton, *Cultivating Open Government Data Platform Ecosystems Through Governance: Lessons from Buenos Aires, Mexico City and Montevideo*, 37 GOV'T. INFO. Q. July 2020, at 2 (commenting, after review of open data experiences across the Americas, that "after [a] decade of open data initiatives, few, if any, economic and social benefits have been achieved") (citation omitted).

operate open data portals or aspire to create them. The recommendations prioritize making government data not just accurate, but actionable—equipping the press and public with information in a way that empowers civic changemaking. Finally, Part V concludes by emphasizing why the open data movement has such promise to transform the way people interact with government, and how news organizations have an essential role to play in making sure the “information revolution” is not tainted by misinformation or disinformation.

I. THE STRUGGLES (AND POTENTIAL) OF LOCAL JOURNALISM

Across the United States, fewer and fewer communities are served by well-staffed daily newspapers keeping watch over civic affairs.⁹ According to an analysis by the Pew Research Center released April 2020, between 2008 and 2019 the number of newspaper newsroom employees dropped by 51 percent.¹⁰ This analysis, the Center notes, does not factor in the economic effects brought on by the novel coronavirus pandemic, which led to even more newsroom layoffs and even closures.¹¹ The drop in newsroom employment follows from diminishing revenue, as readers cease purchasing print newspapers and advertisers find other online outlets to target consumers, chiefly Google and Facebook.¹² The transition from paper to digital journalism has been rocky for local newsrooms, with advertisers unwilling to pay the same rate for online placement that print once commanded.¹³ Declining revenue felled once-acclaimed

⁹ Clara Hendrickson, *Local Journalism in Crisis: Why America Must Revive its Local Newsrooms*, BROOKINGS INST. (Nov. 12, 2019), <https://www.brookings.edu/research/local-journalism-in-crisis-why-america-must-revive-its-local-newsrooms/> (citing research showing that more than 65 million Americans live in counties served by only one newspaper or none at all).

¹⁰ Mason Walker, *U.S. Newsroom Employment has Fallen 26% Since 2008*, PEW RSCH. CTR. (July 13, 2021), <https://www.pewresearch.org/fact-tank/2021/07/13/u-s-newsroom-employment-has-fallen-26-since-2008/>.

¹¹ Gabby Miller, *More than 6,150 news workers were laid off amid the COVID-19 pandemic*, COLUM. JOURNALISM REV. (Dec. 10, 2021), https://www.cjr.org/tow_center_reports/more-than-6150-news-workers-laid-off.php.

¹² See Helen Coster, *Google, Facebook Pledged Millions for Local News. Was it Enough?*, REUTERS (June 10, 2021, 3:50 AM), <https://www.reuters.com/business/media-telecom/google-facebook-pledged-millions-local-news-was-it-enough-2021-06-10/> (reporting that “Google and Facebook accounted for 54% of U.S. digital advertising revenue in 2020”).

¹³ See Andrey Mir, *The Press Now Depends on Readers for Revenue and That’s a Big Problem for Journalism*, DISCOURSE (July 28, 2021),

daily newspapers in major markets including Denver and Seattle, while newspapers elsewhere ceased printing and delivering papers on a daily basis to save money.¹⁴

One of the few bright spots in the news industry has been the emergence of data journalism as a discipline, bringing high-end coding expertise into newsrooms that enables reporters to analyze and use previously unmanageable databases.¹⁵ One measure of this expanding discipline is participation in training opportunities offered by the nonprofit National Institute for Computer Assisted Reporting, which drew more than 1000 people to its 2018 national conference after years of attracting between 300 and 500.¹⁶ In one notable example, Stanford University's "Big Local News" project is tapping into the expertise of Stanford computer scientists, in partnership with newsrooms across the country, to examine vast sets of data about the workings of the criminal justice system, including the demographics of motorists pulled over by police.¹⁷ Data-driven journalism—which lends itself to visual depictions as opposed to vast blocks of text—plays to the interests of an easily distracted

<https://www.discoursemagazine.com/culture-and-society/2021/07/28/the-press-now-depends-on-readers-for-revenue-and-thats-a-big-problem-for-journalism/> ("Publications sold digital ads but couldn't charge anywhere near what they had charged for print ads. As the saying went, print advertising dollars were replaced by digital dimes and then by mobile pennies, as smartphones became the dominant method for accessing the internet.").

¹⁴ See Joanna Penn, *How do Newspapers Affect Civic Life? Data and Analysis on Seattle and Denver*, JOURNALIST'S RES. (Mar. 11, 2014), <https://journalistsresource.org/politics-and-government/how-do-newspapers-affect-civic-life-data-analysis-seattle-denver/> (reporting that researchers documented decline in civic engagement when Seattle's *Post-Intelligencer* and Denver's *Rocky Mountain News* ceased publishing); see also Elahe Azadi, *Newspapers Keep Eliminating Print Days. They Say it's for the Best.*, WASH. POST (Apr. 12, 2022, 7:00 AM), <https://www.washingtonpost.com/media/2022/04/12/gannett-newspaper-print-days/> (reporting that Gannett, America's largest newspaper chain, eliminated a day of print publication at its 136 newspaper nationwide, part of growing trend in which newspapers in Cleveland, Tampa and elsewhere ceased delivering newspapers daily).

¹⁵ See Emilia Ruzicka, *The Rise of Data Journalism*, BROWN UNIV. DATA SCI. INITIATIVE (July 15, 2021), <https://www.brown.edu/initiatives/data-science/news/2021/07/rise-data-journalism> (stating that "the modern field of data journalism exploded in 2008" with the launch of data-focused news sites such as FiveThirtyEight).

¹⁶ Samata Joshi, *NICAR 2018 Attracts More Than A Thousand Data Journalists in Chicago*, MEDIUM (Mar. 17, 2018), <https://medium.com/samata-joshi/nicar-2018-attracts-more-than-a-thousand-data-journalists-in-chicago-444cb7ff57b9>.

¹⁷ Cheryl Phillips, *Uncovering the Truth: Exploring the Benefits of Federated Databases for Policing Records*, DATAJOURNALISM.COM (Dec. 6, 2022), <https://datajournalism.com/read/longreads/database-police-records-accountability>.

online audience that is accustomed to consuming information in bite-sized visualizations.¹⁸ Data journalism meets the needs and interests of a generation of readers raised on smartphones and social discussion forums for interactivity with the information they consume, which in turn deepens the news organization's engagement with its audience.¹⁹ By turning data into visualizations, journalists can make information appealing to the eye and more easily consumed by on-the-go readers scrolling on their phone screens. Publishing data online allows news organizations to offer regularly refreshed editorial content that grows over time as the data evolves, as was evident with data-driven coverage of the spread of COVID-19 during the coronavirus pandemic.²⁰

News organizations have embraced data journalism as a way to "make journalism more systematic, accurate, and trustworthy."²¹ Public-opinion polls confirm the urgency of repairing strained trust between the public and the press. According to a 2020 poll of 20,000 adults conducted by Gallup and the Knight Foundation, 83 percent of Americans perceive "a fair amount" or "a great deal" of political bias tainting news coverage, and 68 percent regard bias as a "major" problem, with self-identified conservatives far more likely to express distrust and to believe that the demonization of news organizations is justified.²² Presenting information in the form of hard numbers that can be tracked and validated can be at least a partial antidote to what some commentators have called factual plurality. Factual plurality is the lack of an

¹⁸ See Dream McClinton, *Global Attention Span is Narrowing and Trends Don't Last as Long, Study Reveals*, GUARDIAN (Apr. 17, 2019, 12:32 AM), <https://www.theguardian.com/society/2019/apr/16/got-a-minute-global-attention-span-is-narrowing-study-reveals> (reporting that researchers have documented diminishing attention spans worldwide as audiences become overwhelmed with bombardment of new online content).

¹⁹ Jan Lauren Boyles & Eric Meyer, *Letting the Data Speak: Role Perceptions of Data Journalists in Fostering Democratic Conversation*, 4 DIGIT. JOURNALISM, no. 7, 2016, at 944.

²⁰ See, e.g., Cedar Rapids Gazette, *New Iowa COVID Data Website Built by Iowa Newspapers Shares Coronavirus Hospitalizations, Deaths, Cases*, GAZETTE <https://www.thegazette.com/health-care-medicine/new-iowa-covid-data-website-iowa-newspapers-coronavirus-hospitalizations-deaths-cases/> (Feb. 24, 2022, 12:00 PM) (reporting that collaborative of Iowa newspapers banded together to publish data about COVID-19 hospitalizations and other pandemic metrics after state Department of Public Health discontinued its standalone COVID-19 data site).

²¹ Rodrigo Zamith, *Transparency, Interactivity, Diversity, and Information Provenance in Everyday Data Journalism*, 7 DIGIT. JOURNALISM, no. 4, 2019, at 470.

²² KNIGHT FOUN., *American Views 2020: Trust, Media and Democracy*, KNIGHT FOUN. (Apr. 20, 2020) <https://knightfoundation.org/reports/american-views-2020-trust-media-and-democracy/>.

agreed-upon baseline of reality on which public-policy positions can be based.²³

The imperative for journalists to “show their work” has only heightened with the proliferation of deceitful websites masquerading as community news purveyors, operated by partisan actors disguising opinion as fact, a practice that has become known in shorthand as “pink slime news.”²⁴ In other words, people are not just more skeptical, but more *justifiably* skeptical because of widespread efforts by malign actors to deceive them for financial or political gain. At the same time, journalists are being widely urged to abandon traditional concepts of “both-sides” neutrality, and to cut through bad-faith misrepresentations by getting at the underlying facts.²⁵ By basing news reports on quantifiable data—and showing where journalists’ conclusions actually come from—news organizations can better combat the challenges of skepticism, distraction, and disinformation. In the context of journalism, transparency is understood to encompass both disclosure transparency showing how news is selected and gathered, and participatory transparency involving the audience in the process of formulating and prioritizing news.²⁶ Both of these elements can create a greater sense of legitimacy from the perspective of audiences, and both are aided by the ability to fortify news reporting with public-facing data sets. Some news organizations have even invited the public to contribute to, or help analyze, large data sets in a form of crowdsourcing, thus turning journalism from a spectator sport into a participatory sport.²⁷ By not just

²³ Nikki Usher, *Re-Thinking Trust in the News: A Material Approach Through “Objects of Journalism”*, 19 JOURNALISM STUDIES, no. 4, 2018, at 564, 572.

²⁴ See Priyanjana Bengani, *As Election Looms, a Network of Mysterious ‘Pink Slime’ Local News Outlets Nearly Triples in Size*, COLUM. JOURNALISM REV. (Aug. 4, 2020), <https://www.cjr.org/analysis/as-election-looms-a-network-of-mysterious-pink-slime-local-news-outlets-nearly-triples-in-size.php> (describing proliferation of more than 1,200 documented websites nationwide financed by conservative donors to promote partisan talking points while masquerading as ordinary news outlets).

²⁵ See Margaret Sullivan, *This Week Should Put the Nail in the Coffin for ‘Both Sides’ Journalism*, WASH. POST (Aug. 16, 2017), [https://www.washingtonpost.com/lifestyle/style/this-week-should-put-the-nail-in-the-coffin-for-both-sides-journalism/2017/08/16/77c6668a-8292-11e7-b359-](https://www.washingtonpost.com/lifestyle/style/this-week-should-put-the-nail-in-the-coffin-for-both-sides-journalism/2017/08/16/77c6668a-8292-11e7-b359-15a3617c767b_story.html)

[15a3617c767b_story.html](https://www.washingtonpost.com/lifestyle/style/this-week-should-put-the-nail-in-the-coffin-for-both-sides-journalism/2017/08/16/77c6668a-8292-11e7-b359-15a3617c767b_story.html) (drawing on then-President Donald Trump’s speech reacting to white supremacist march in Charlottesville, Virginia, in which he appeared to blame civil-rights protesters equally for ensuing violence, to call for news organizations to abandon 50/50 balance when facts are clearly on one side of debate: “The best way to be fair is not to be falsely evenhanded, giving equal weight to unequal sides. It’s to push for the truth, and tell it both accurately and powerfully.”).

²⁶ Zamith, *supra* note 20.

²⁷ See Amanda Zamora, *Crowdsourcing Campaign Spending: What We Learned From Free the Files*, PROPUBLICA (Dec. 12, 2012),

identifying the origin of facts cited in news stories but actually linking to or embedding tables and spreadsheets in their reporting, journalists can more effectively respond to skeptics, or opportunistic politicians, who question the legitimacy of unwelcome disclosures in news stories.²⁸

II. THE STATE OF OPEN DATA

A. Access Law 101

For decades, news organizations and concerned citizens have depended on a pull model of getting information from government agencies. In such models, the requester identifies a document that memorializes the desired data, and the agency with the document—sometimes after great delay, expense, and resistance—produces it.²⁹ Freedom of information or public records laws date back to the early 20th century, but the movement toward liberating government documents was supercharged by the 1966 enactment of the Freedom of Information Act (FOIA), a federal statute that became a template for state-level lookalikes.³⁰ Now, every U.S. state and territory has a

<https://www.propublica.org/article/crowdsourcing-campaign-spending-what-we-learned-from-free-the-files> (describing results of crowdsourcing project by online investigative nonprofit ProPublica, which drew on assistance from nearly 1,000 reader volunteers to review and catalog entries in Federal Communications Commission database of campaign advertising buys on FCC-regulated broadcast stations in 33 key geographic markets).

²⁸ See Amy Mitchell et al., *Many Americans Say Made-Up News is a Critical Problem That Needs to be Fixed*, PEW RSCH. CTR. (June 5, 2019), <https://www.pewresearch.org/journalism/2019/06/05/many-americans-say-made-up-news-is-a-critical-problem-that-needs-to-be-fixed/> (reporting results of a 2019 survey of 6,127 U.S. adults, 50 percent of whom identified “made-up” news and information as a major problem for the United States, about the same percentage as named wealth inequity, violent crime and climate change).

²⁹ See Christina Koningsor, *Transparency Deserts*, 114 NW. U. L. REV. 1461, 1472–73 (2020) (describing chronic understaffing in federal agencies that delays FOIA fulfillment: “Members of Congress routinely accuse executive officials of creating obtuse procedures deliberately designed to confuse requesters. And members of the media argue that enumerated exemptions have been construed so broadly that they swallow vast categories of records and that excessive time delays make the statute nearly unusable.”) (footnotes omitted); see also Jan Whittington et al., *Push, Pull, and Spill: a Transdisciplinary Case Study in Municipal Open Government*, 30 BERKELEY TECH. L.J. 1899, 1904 (2015) (describing distinction between “push” model of affirmative publication of databases versus “pull” of requesters demanding public records).

³⁰ See Adriana S. Cordis & Patrick L. Warren, *Sunshine as Disinfectant: The Effect of State Freedom of Information Act Laws on Public Corruption*, 115 J. PUB. ECON. 18, 23 (2014) (“Progress on guaranteeing access to information... was relatively limited until the 1970s. In the last 40 years,

facsimile of FOIA.³¹ These laws are widely recognized as having all manners of salutary civic benefits. They act as a check on government overreach and corruption and provide the public with insight into the workings of the government to inform better quality participation in government decision-making.³²

FOIA laws have proven to be powerful tools enabling journalists to tell well-supported stories about the failings of government agencies and officials.³³ While stories powered by federal FOIA reliably win national awards and permeate the public consciousness, each day these statutes are used by workaday reporters documenting routine activities of local police departments, school boards, and city or county commissioners.³⁴ Journalists are not the only or even the main users of public-records laws. A 2017 study found that just 7.6 percent of federal FOIA requests were filed by news organizations, with businesses, law firms, and nonprofits all filing substantially more.³⁵ Environmentalists, consumer

most states enacted new open record statutes, amended existing statutes, or rewrote their statutes in an effort to strengthen the laws, often to clarify or broaden their scope...”).

³¹ Jonathan Anderson et al., *Policy Liberalism and Public Records Laws in the American States*, 27 COMM. L. & POL’Y 30, 41 (2022).

³² See Altman et al., *supra* note 6, at 1986 (“Public records help members of the public, including journalists, learn about criminal and police activity in their neighborhoods, investigate the prevalence of public safety issues they encounter, and advocate reforms based on the patterns they discover.”) (footnote omitted); see also Oren Perez, *Open Government, Technological Innovation, and the Politics of Democratic Disillusionment: (E)-Democracy from Socrates to Obama*, 9 I/S: J.L. & POL’Y FOR INFO. SOC’Y 61, 72 (2013) (explaining that citizen participation in government is understood to provide a “collective wisdom” benefit by drawing on the public’s knowledge as well as a “check on the bureaucratic process”).

³³ See Koningisor, *supra* note 28, at 1467 (stating that “many reporters themselves view state public records statutes as deeply integral to the reporting process”).

³⁴ See Paul Aker, *Towards Darkness: Ohio’s Presumption of Openness Under the Public Records Act*, 41 CAP. U. L. REV. 361, 367 (2013) (“In a news world crowded by public information officers, press releases, and spin-doctors, reporters rely on public documents to expose the unvarnished facts about government activity. Amid that backdrop, public records produce enlightening stories about everything from government officials taking millions of dollars in illegal overtime payments, to contaminated state DNA labs, to corruption in the Ohio Attorney General’s Office.”) (footnotes omitted); see also Brooke Barnett, *Use of Public Record Databases in Newspaper and Television Newsrooms*, 53 FED. COMM. L.J., no.3, at 557, 565 (2001) (reporting results of survey of local television stations and newspapers, which found that 55.9 percent reported that they consult public records for investigative stories “most of the time” or “often”).

³⁵ Cory Schouten, *Who Files the Most FOIA Requests? It’s Not Who You Think.*, COLUM. JOURNALISM. REV. (Mar. 17, 2017), <https://www.cjr.org/analysis/foia-report-media-journalists-business-mapper.php>.

advocates, and other public-interest watchdogs have been especially avid and effective users of FOIA, obtaining and publishing records that help lend credibility to their advocacy.³⁶ But, as effective as they can be, FOIA laws also have well-documented shortcomings.

Both the federal statute and its state analogs are riddled with easily abused exemptions, such as one excusing production of documents that might shed light on the “deliberative process” by which agencies formulate public policy.³⁷ Judges typically hesitate to second-guess agencies’ decisions to categorize documents as exempt, particularly where the agencies invoke national security, public safety, or medical privacy as their reason for exemption, no matter how attenuated the connection to the requested information.³⁸

In recognition that the public should not be forced to demand the same documents over and over through repeated cycles of the FOIA process, Congress modernized the federal statute in 1996 to require some degree of affirmative disclosure.³⁹ Under the Electronic Freedom of Information Act Amendments of 1996 (commonly known as “E-FOIA”), agencies are directed to publish in electronic reading rooms

³⁶ See Altman, et al., *supra* note 6, at 1979–80 (describing instances in which advocates used federal data about highway accidents and doctors’ prescribing practices to generate public awareness for their causes).

³⁷ See 5 U.S.C. § 552(b)(5) (providing that agencies may withhold “inter-agency or intra-agency memorandums or letters that would not be available by law to a party other than an agency in litigation with the agency,” an exemption widely referred to as the deliberative process exemption). A number of states have comparable – or even broader – exemptions, such as Michigan, where agencies may withhold otherwise-public records of communications “within a public body or between public bodies of an advisory nature to the extent that they cover other than purely factual materials and are preliminary to a final agency determination of policy or action” if the agency shows that its interest in promoting “frank communication ... clearly outweighs the public interest in disclosure.” MICH. COMB. LAWS § 15.243(1)(m). See also Edward J. Imwinkelried, *The Government’s Increasing Reliance on—and Abuse of—the Deliberate Process Evidentiary Privilege: “[T]he Last Will Be First*”, 83 MISS. L.J. 509, 526–28 (2014) (critiquing judicial expansion of scope of what is considered “deliberative” material exempt from production under both evidence law and FOIA law).

³⁸ See Barry Sullivan, *FOIA and the First Amendment: Representative Democracy and the People’s Elusive “Right to Know”*, 72 MD. L. REV. 1, 84 (2012) (“Government officials routinely produce vague and conclusory affidavits to justify the withholding of information, and courts routinely find those affidavits sufficient to justify the withholding.”)(footnote omitted).

³⁹ See Michael Herz, *Law Lags Behind: FOIA and Affirmative Disclosure of Information*, 7 CARDOZO PUB. L. POL’Y & ETHICS J. 577, 587–88 (2009) (describing how Congress mandated creation of online reading rooms for oft-requested federal documents, but noting that many agencies have failed to update the reading rooms so that the legislation “has not wholly lived up to its promise”).

any documents or records processed and disclosed in response to a FOIA request that “the agency determines have become or are likely to become the subject of subsequent requests for substantially the same records.”⁴⁰ These amendments, in theory, were intended to make information more easily accessible for journalists, watchdogs, and other members of the general public.

But even the best-written and most rigorously enforced FOIA laws entitle the public to see only records that government agencies have already created. There is no duty for an agency to create a compilation to satisfy a FOIA request.⁴¹ In other words, FOIA can make existing data accessible, but it cannot compel agencies to actually gather and assemble data in an intelligible way. For that, the public must either depend on a state open data mandate or must hope that agencies are sufficiently well-managed and inquisitive that they will compile high-value data sets of their own volition.

The interplay between FOIA and open data is—or should be—readily apparent: The more information that government agencies proactively make public, the less time and effort they should have to spend fulfilling requests from people who need the information.⁴² Indeed, government agencies have already reported that open data saves them money in FOIA fulfillment and reduces habitual backlogs in FOIA requests, producing faster results for requesters.⁴³ A study by the government watchdog organization Reinvent

⁴⁰ 5 U.S.C. § 552(a)(2)(D).

⁴¹ See *Ochsner v. N.C. Dept. of Revenue*, 835 S.E.2d 491, 502 (N.C. App. 2019) (“The Public Records Act does not require an agency to create or compile records responsive to a request if those records do not exist; the agency must produce only the records which already exist.”) (citation omitted); *Faulkner v. Washington Dept. of Corrections*, 332 P.3d 1136, 1140 (Wash. App. 2014) (“An agency does not have a duty to produce a record that does not exist.”) (citation omitted).

⁴² Bev Wilson & Cong Cong, *Beyond the Supply Side: Use and Impact of Municipal Open Data in the U.S.*, 58 TELEMATICS & INFORMATICS, no. 101526, 2021, at 4 (citing potential benefit of “minimizing redundancies in responding to requests for public information”).

⁴³ See Alena Stern, *Research: Cities can Save Time on Records Requests by Doing Open Data Right*, SUNLIGHT FOUND. (Oct. 9, 2018, 3:28 PM), <https://sunlightfoundation.com/2018/10/09/research-cities-save-time-on-records-requests-by-doing-open-data-right/> (reporting on results of study of FOIA requests processed by 52 municipalities, which found that municipalities with well-populated open data websites saw 33 percent decrease in volume of requests for records compared with a control group that lacked robust data portals); see also Wilson & Cong, *supra* note 41, at 8 (reporting results of authors’ interviews with local government data custodians: “When as much data is released as possible on the open data portal, they either see a significant reduce of FOIA requests, or requests that are more specific and slightly different than what has been released as open datasets.”).

Albany suggests that by looking at patterns of FOIA requests, agencies can identify the highest-demand records that belong on open data sites, making access more efficient for both requesters and custodians.⁴⁴ Thus, affirmatively producing information on open data sites is not just convenient for requesters, but also can be advantageous to government agencies as well.

B. The Quest to Liberate Data

Open data is not just a technical term, but a multi-decade movement encompassing the goals of the open source and open government movement.⁴⁵ Open Knowledge Foundation, an influential British-based nonprofit, has developed a guidebook for stewards of data setting forth guiding principles about what it means for data to be truly open, including: availability for free or at reasonable cost, without restriction on re-using or re-mixing it, and with no discrimination against users (i.e., that full access must extend to all users, even commercial ones).⁴⁶ Making government data more readily available to an external audience is understood to have multiple benefits: It can fuel private-sector innovation and growth, promote better-quality civic participation, and help agencies themselves operate more efficiently.⁴⁷

Almost as long as the World Wide Web has existed, government agencies have published information online to public audiences. But a website is not truly an open data platform just because it contains information. For instance, a municipality might post a list of polling places or library branches, but that list would not be classically regarded as open data.⁴⁸ The phrase open data instead suggests information that might be possible *not* to publish, which plainly would not be the case with data such as the location

⁴⁴ *FOIL Request Logs as Open Data Sensor*, REINVENT ALBANY (Aug. 13, 2014), <https://reinventalbany.org/2014/08/using-foil-logs-to-guide-the-publication-of-open-data/>.

⁴⁵ See Noveck, *supra* note 5, at 10–12 (tracing open government data movement's origins to early 1940s and linking it with broader concepts about openness of scientific data).

⁴⁶ Open Knowledge Found., *What is Open Data?*, OPEN DATA HANDBOOK, <https://opendatahandbook.org/guide/en/what-is-open-data/>.

⁴⁷ Frederik Zuiderveen Borgesius et al., *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*, 30 BERKELEY TECH. L.J. 2073, 2078 (2015).

⁴⁸ See Harlan Yu & David G. Robinson, *The New Ambiguity of "Open Government"*, 59 UCLA L. REV. DISC. 178, 189 (2012) (contemplating meaning of "open data" and indicating that it "suggests participation and engagement" such that people might use the data to "reach their own conclusions").

of polling places. For example, a bus schedule is not what is meant by open data, but a spreadsheet of bus routes showing what percentage of buses habitually arrive late is. Data that is purely functional (“here is the deadline for dropping off ballots,” or “here is when the bus will arrive”) does not further the purposes of the open data movement: It does not promote a deeper understanding of how government works, enable a user to draw comparisons and track trends over time, or increase government accountability.

The U.S. federal government’s first open data website, Data.gov, was launched in May 2009.⁴⁹ But nonprofits and advocacy organizations have been building websites to share government-created data for far longer, including the campaign finance reform organization Open Secrets, which launched a searchable database of candidates’ financial disclosure forms in 1998.⁵⁰ The push for openness has come largely from nongovernmental advocacy organizations, such as the Sunlight Foundation and Code for America, which have sought to nudge sometimes-recalcitrant government agencies to adopt the open data mentality. More recent entries into the field, such as Data For Black Lives (focusing on police misconduct)⁵¹ and the Center for Tech and Civic Life (focusing on voter turnout and election integrity)⁵² have taken up advocacy for transparency not as an end in itself, but as a means toward fueling their underlying public policy goals.

The OPEN Government Data Act was passed in January 2019 and requires that federal agencies publish their online information as open data in a standardized format.⁵³ All fifty states have some form of state-level open data portal.⁵⁴ But as of the beginning of 2022, only sixteen states implemented a statutory requirement similar to the OPEN Government Data Act that requires agencies to publish data in a centralized and readable format, meaning that in most states, open data is a matter of agency discretion.⁵⁵

⁴⁹ Hyon Kim, *Data.gov at Ten and the OPEN Government Data Act*, DATA.GOV (May 31, 2019), <https://data.gov/meta/data-gov-at-ten-and-the-open-government-data-act/index.html>.

⁵⁰ Yu & Robinson, *supra* note 47, at 192.

⁵¹ DATA FOR BLACK LIVES, <https://d4bl.org/>.

⁵² CTR. FOR TECH AND CIVIC LIFE, <https://www.techandciviclife.org/>.

⁵³ Tajha Chappellet-Lanier, *The OPEN Government Data Act is Now Law*, FEDSCOOP (Jan. 15, 2019), <https://www.fedscoop.com/open-government-data-act-law/>.

⁵⁴ Meta S. Brown, *States Offer Information Resources: 50+ Open Data Portals*, FORBES (Apr. 30, 2018, 12:07 AM), <https://www.forbes.com/sites/metabrown/2018/04/30/us-states-offer-information-resources-50-open-data-portals/?sh=65fdc8645225>.

⁵⁵ *Research and Policy*, NAT’L CONF. OF STATE LEGISLATURES, (Jan. 25, 2022), <https://www.ncsl.org/research/telecommunications-and-information-technology/state-open-data-laws-and-policies.aspx>.

Predictably, this lack of statutory compulsion means that state and local-level data portals will vary in quality, completeness, and currentness.

Whether because of legal compulsion or because of their own self-interests, government agencies unquestionably have begun harnessing the power and bandwidth of the internet to make data and documents more accessible. Some data are purely statistical—such as active business licenses in a community or logs of 911 emergency calls—and others are more evaluative—such as responsiveness to public complaints about potholes or progress reports on achieving energy conservation goals.⁵⁶ Some reflect inwardly on the performance of government—such as vendor contracts or “open checkbooks”—and others reflect outwardly on citizen interactions with government—such as surveys reflecting average commute times.⁵⁷ Increasingly, government entities are purchasing turnkey open data platforms offered by vendors such as Socrata and CKAN, making it easier for agencies lacking in-house programming expertise to take part in the open data movement.⁵⁸ Still, there is little standardization across levels of government and locations, nor is there any systematic auditing mechanism to verify whether the data being produced is complete and accurate.

Simply creating a data portal, of course, is not the only critical decision point for a government entity. Choices must also be made about which data sets to prioritize and how to display them. As to the method of display, the OPEN Government Data Act specifies that the data published to Data.gov must be in a standardized, machine-readable format.⁵⁹ For example, if a data set is embedded in a map or graph on an open data website, it should also be provided in a raw data format so that users can download it and use it for their own projects. Data.gov also enforces a metadata standard; that is, a standard way of presenting the data where a set of required fields (title, description, tags, last update, etc.) must be filled. Additionally, there should be labels indicating what file types the data set is available in. These are important standards to have at the local level because they offer consistency in how data is presented.⁶⁰ But in some

⁵⁶ See Whittington et al., *supra* note 28, at 1940 (cataloguing some of the 200-plus data sets published through the City of Seattle’s open data portal).

⁵⁷ *Id.*

⁵⁸ David Raths, *When Publishing Open Data, Cities and States Have Variety of Platform Choices*, GOV’T. TECH. (Mar. 2016), <https://www.govtech.com/gov-experience/when-publishing-open-data-cities-and-states-have-variety-of-platform-choices.html>.

⁵⁹ Kim, *supra* note 48.

⁶⁰ See Barbara L. Cohn, *Data Governance: a Quality Imperative in the Era of Big Data, Open Data, and Beyond*, 10 I/S: J.L. & POL’Y FOR INFO. SOC’Y 811, 821

places, agency policies continue working at cross-purposes with the open data movement. In South Carolina, for instance, civil liberties groups are suing to challenge the state court system’s policy forbidding bulk “scraping” of court data from a publicly accessible website, which means researchers and journalists are limited to gathering case-by-case data rather than building their own aggregate databases.⁶¹

C. The Use Case: Open Data as a Journalistic “Force Multiplier”

When 32-year-old Tony Timpa died suddenly outside a Dallas bar, a reporter for the *Dallas Morning News* faced a wall of secrecy from the local police department, which offered contradictory accounts of the circumstances of the death.⁶² Although police made official reports about the death invisible, reporter Cary Aspinwall was able to match the location of the death to an unnamed entry in the police department’s online data portal, which gave her the information she needed to pursue the police incident report and body-cam video.⁶³ That video footage—released only after a prolonged legal battle—showed that Timpa did not die of a heart attack or seizure, as police initially said, but rather was pinned down by police and immobilized until he lost consciousness, in a scenario eerily reminiscent of the nationally publicized killing of George Floyd in Minneapolis.⁶⁴ While Aspinwall might have broken the Timpa story anyway eventually, the existence of a public-facing database of 911 calls undoubtedly made her work faster and easier; had the database not existed, she would have been forced to try to pry the 911 call data loose from the police department under the Texas Public Information Act and then comb through it for a

(2015) (making the point that interoperability of data among different users requires a baseline metadata scheme, data dictionary, and other standardized protocols, so users can understand what they are getting).

⁶¹ Gillian Diebold, *South Carolina’s Misguided Restrictions on Scraping Judicial Data*, CTR. FOR DATA INNOVATION (Mar. 30, 2022), <https://datainnovation.org/2022/03/south-carolinas-misguided-restrictions-on-scraping-judicial-data/>.

⁶² Cassandra Jaramillo, *How the Dallas Police’s Open Portal Led to the True, Untold Story of Tony Timpa’s Death*, DALLAS MORN. NEWS (Nov. 17, 2021, 1:38 PM), <https://www.dallasnews.com/news/investigations/2021/11/17/how-the-dallas-polices-open-portal-led-to-the-true-untold-story-of-tony-timpas-death/>.

⁶³ *Id.*

⁶⁴ *Id.*

call matching Timpa's, with all of the customary cost and delay that typically accompany public records requests.⁶⁵

As Aspinwall did in Dallas, journalists have regularly found ways to use open data portals to more effectively report news. News outlets in New York often draw on data published online by municipal agencies to report on trends in homelessness, crime, and other metrics of civic well-being.⁶⁶ Small, thinly staffed news organizations have benefited from drawing on open data to tell stories they might have lacked the capacity to tell on their own; for instance, Streetsblog NYC, which lists a staff of six people on its website, used data from the city's transportation department to show that Mayor Eric Adams was significantly behind schedule on promised roadway improvements to protect pedestrians and cyclists.⁶⁷

The *Staten Island Advance* newspaper has avidly used New York City's open data platform to report on local affairs in the borough, including regular updates on where motorists are incurring the most speeding tickets by way of automated speed cameras.⁶⁸ The newspaper's use of speed-cam data illustrates both the potential and the shortcomings of the current state of municipal open data: Reporters pointed out that because the data produced by the city is not broken down by county, there is no easy way for the public to fact-check city officials' claim that 40 percent of all tickets go to nonresidents of New York City.⁶⁹ The newspaper's coverage is not merely a matter of idle curiosity: Reporters have tapped into city-produced data to explore the role of distracted

⁶⁵ *Id.* (describing Dallas police records as "difficult to obtain").

⁶⁶ See, e.g., Chau Lam, *Number of Homeless Adults in NYC Shelters Sets New Record by the Day*, *GOTHAMIST* (Oct. 6, 2022), <https://gothamist.com/news/number-of-homeless-adults-in-nyc-shelters-sets-new-record-by-the-day>.

⁶⁷ Gersh Kuntzman, *Update: 'Get Stuff Done' Mayor Far Behind on His Vow to Fix 1,000 Intersections*, *STREETSBLOG NYC* (Sept. 21, 2022), <https://nyc.streetsblog.org/2022/09/21/promises-promises-get-stuff-done-mayor-far-behind-on-his-vow-to-fix-1000-intersections/>.

⁶⁸ Erik Bascome, *S.I. Speed Cameras Net Almost \$10M Since June: Here Are the Most Lucrative Locations*, *SILIVE.COM* (Feb. 21, 2022, 12:17 PM), <https://www.silive.com/news/2022/02/si-speed-cameras-net-almost-10m-since-june-here-are-the-most-lucrative-locations.html>.

⁶⁹ Erik Bascome, *NYC Speed Cameras: Cars Registered Outside the City Account For Over 40% of Violations, DOT Says*, *SILIVE.COM* (Mar. 10, 2022, 85:30 AM), <https://www.silive.com/news/2022/03/nyc-speed-cameras-cars-registered-outside-the-city-account-for-over-40-of-violations-dot-says.html>.

driving in car crashes⁷⁰ and to document substandard conditions in government-subsidized rental housing.⁷¹

An archetypal open data success story came in 2016 when a lone blogger operating a government watchdog website, I Quant NY, drew on open data to show that police were regularly issuing parking tickets to legally parked cars.⁷² Ben Wellington, a trained statistician, noticed that police were ticketing cars in his neighborhood for parking in front of pedestrian ramps, even though they were legally parked.⁷³ He then used the New York City open data portal to discover that the practice was widespread, amounting to an estimated \$1.7 million in tickets per year that might have been wrongfully issued.⁷⁴ When Wellington brought his findings to the attention of the New York Police Department, the agency acknowledged the mistake, and agreed to retrain officers and electronically monitor their ticketing practices to check for irregularities.⁷⁵

News organizations are not, of course, the only beneficiaries of readily accessible government data. Advocacy organizations, which are also heavy users of FOIA laws, have found a powerful tool in the data published by municipalities. For example, the ACLU of South Carolina used the Charleston Police Department's open data portal to analyze patterns of arrests and citations, concluding that police spent most of their time arresting and citing people for low-level nonviolent offenses like marijuana possession, the impact of which fell disproportionately on Black people.⁷⁶

⁷⁰ See e.g., Gail Lubin, *Here is What's to Blame for Staten Island Crashes that Resulted in Injuries*, SILIVE.COM (Apr. 18, 2021, 2:20 PM), <https://www.silive.com/news/2021/04/here-is-whats-to-blame-for-staten-island-crashes-that-resulted-in-injuries.html>.

⁷¹ See, e.g., Giavanni Alves, *One Park Hill Building Tops NYC List of Staten Island Apartments with Most Violations, Report Says*, SILIVE.COM (Mar. 23, 2022, 4:36 PM), <https://www.silive.com/news/2022/03/one-park-hill-building-tops-nyc-list-of-staten-island-apartments-with-most-violations-report-says.html>.

⁷² Chris Baraniuk, *New York Blogger Reveals Parking Ticket Errors*, BBC (May 12, 2016), <https://www.bbc.com/news/technology-36275930>.

⁷³ *Id.*

⁷⁴ Ben Wellington, *The NYPD Was Systematically Ticketing Legally Parked Cars for Millions of Dollars a Year- Open Data Just Put an End to It*, IQUANT NY (May 11, 2016), <https://iquantny.tumblr.com/post/144197004989/the-nypd-was-systematically-ticketing-legally>.

⁷⁵ *Id.*

⁷⁶ Tim Renaud, *ACLU of South Carolina Released Report They Said Shows "Staggering Racial Disparities" in Charleston Arrests, Citations*, COUNT ON NEWS 2 (Oct. 27, 2021, 6:34 PM), <https://www.counton2.com/news/local-news/aclu-of-south-carolina-released-report-they-said-shows-staggering-racial-disparities-in-charleston-arrests-citations/>.

Journalists and researchers might understandably have a love/hate relationship with government data portals. When a requester asks a government agency for public records, the requester expects to receive just that: records. If the agency instead sends the requester to an open data platform for a do-it-yourself FOIA response, that response may not be wholly satisfying. Researchers from the University of Wisconsin, who were testing the accessibility of common data sets from higher education institutions, encountered this issue when the University of Nebraska declined to fulfill their FOIA request and simply referred them to the state’s transparency website—which, the researchers found, did not allow university-specific queries.⁷⁷ Other universities, the researchers reported, referred them to public-facing websites that only incompletely answered their needs. For instance, some of the sites did not provide functionality for downloading the data, which points to the risk that agencies may structure their data portals to purposefully limit the information’s usefulness.⁷⁸ Moreover, there are genuine concerns that the very existence of open data portals could become a cop-out to avoid confronting the well-documented problems afflicting FOIA compliance, or they could drain resources away from FOIA fulfillment.⁷⁹

As demonstrated by the aforesaid work of reporters at Streetsblog NYC, the *Staten Island Advance* and elsewhere, an open data portal that is rich with contemporary datasets can be a labor-saver for news organizations, particularly small-staffed organizations (or, like Ben Wellington, one-man blogging operations). It is doubtful that a news organization as small as Streetsblog NYC could have invested the time and labor to create its own compilation of the 395 road improvement projects described in its article about the Adams administration, had the compilation not already existed on the city’s open data site. Indeed, in the case of Streetsblog NYC, the open data site was the difference-maker

⁷⁷ Jonathan Anderson & Sarah K. Wiley, *Freedom of the Database: Auditing Access to Structured Data*, 3 J. CIVIC INFO., June 2021, at 30, 42 n.96 .

⁷⁸ See *id.* at 49 (“Public agencies may proactively disclose limited data or post data in restrictive formats to control information.”)

⁷⁹ See Daniel Berliner et al., *The Future of FOIA in an Open Government World: Implications of the Open Government Agenda for Freedom of Information Policy and Implementation*, 63 VILL. L. REV. 867, 888 (2018) (suggesting that implementation of an open data strategy “may actually result in fewer resources available for FOI” and relating that this fear actually materialized in Mexico, when the task of populating open data sites was assigned to the same employees responsible for fulfilling FOI requests); see also *id.* at 884 (stating that “by allowing governments to claim credit for less politically-inconvenient reforms, the broader open government agenda may enable a false sense of transparency even where access-to-information rights are nonexistent or ineffective”).

between a story and no story, because – as the article indicated – the Department of Transportation failed to respond to multiple requests from the news organization for information about the status of the mayor’s promised infrastructure improvements.⁸⁰

News organizations should be able to use data portals as a labor-saver, allowing them to cut out the gathering process and go straight into the analysis. But for the data to be useful, it must pass the test of trustworthiness and relevance, addressing newsworthy public concerns that audiences care about. The first generation of open data platforms has not always reliably passed either of these tests.

D. Police Use-of-Force Data: A Case Study

To assess whether municipalities are predictably putting high-value data sets online, we present findings from a study we conducted with the Brechner Center for Freedom of Information at the University of Florida sampling thirty municipal websites for reports about police force. The thirty municipalities were chosen according to size cohort: the ten most populous cities in the country, ten cities with a population of about 200,000, and ten cities with a population of between 70,000 and 90,000. Researchers utilized both the open data portal for the city police department and the portal for the corresponding host city, seeking three related data points: (1) instances in which police officers used force; (2) of those instances, the subset of cases in which officers used firearms; and (3) complaints from the public concerning the use of force.

Of the ten largest cities (Cohort 1), all included at least one of the three targeted data points on their open data portals, and two—New York City and San Antonio—included all three.⁸¹ Of the ten medium-sized cities (Cohort 2), six did not provide any discernible data about any of the queries at all. Two mid-sized cities—Norfolk, Virginia, and Tacoma, Washington—provided all three data points.⁸² None

⁸⁰ Kuntzman, *supra* note 66.

⁸¹ *Annual Use of Force/Firearms Discharge Report Data Tables*, N.Y. POLICE DEPT., <https://www1.nyc.gov/site/nypd/stats/reports-analysis/firearms-discharge.page>; *NYPD Member of Service Histories*, N.Y. POLICE DEPT., <https://www1.nyc.gov/site/ccrb/policy/MOS-records.page>; *SAPD Open Data Initiative*, SAN ANTONIO POLICE DEPT., <https://www.sanantonio.gov/SAPD/SAPD-Open-Data-Initiative#182284287-critical-incident-procedure>; *Officer Involved Injuries or Death Caused by the Use of a Firearm*, SAN ANTONIO POLICE DEPT., <https://www.sanantonio.gov/SAPD/Officer-Involved-Shooting>.

⁸² *Police Use of Force and Citizen Complaint Incidents*, NORFOLK OPEN DATA, <https://data.norfolk.gov/Public-Safety/Police-Use-of-Force-and-Citizen-Complaint-Incident/fcqe-uvnb>; *Police and Use of Deadly Force in*

of the ten small cities (Cohort 3) had any of the three data points online. In one instance (Trenton, New Jersey), two of the three data points—use of force and officer-involved shootings—were readily findable online through a statewide open data portal maintained by the state attorney general, not by the municipality.

The observation that the sophistication of the data portal correlates with the size of the city makes logical sense. Large cities will have the capacity to hire technologists to maintain open data portals and should be able to afford more elaborate recordkeeping.⁸³ But the distinction speaks to a significant data divide between have and have-not communities. Communities such as New York and Los Angeles not only have greater open data capabilities, but they also have well-funded news operations that can unearth data, while cities such as Edmond, Oklahoma, and Franklin City, Tennessee, may not have either. People in smaller communities thus face an informational double-whammy: Civically valuable data is reliably available neither from the news media nor from the government.

III. ARRESTED DEVELOPMENT: WHY OPEN DATA HAS UNDERACHIEVED

America's first generation of experimentation with gathering and disseminating data sets on a wide scale has, so far, proven more useful and engaging to government insiders than to outsiders.⁸⁴ For those working within government, open data can promote greater sharing of information between sister agencies or even within the same agency, allowing for more efficient operations. For instance, during 2016, California legislators directed the state Department of Water Resources to create a platform integrating data points from all levels of government about the location and quantity

Tacoma, CITY OF TACOMA,
https://www.cityoftacoma.org/government/city_departments/City_Managers_Office/police_and_use_of_deadly_force_in_tacoma.

⁸³ See What Works Cities, (clarify more clearly what this is? A organization? Online resource?) certifies communities for excellence in producing data, noted in a January 2021 report that “in general, the larger the city’s population, the higher the Certification score”. John Collier & Lauren Su, *Data-Driven Government in Large & Small Cities*, MEDIUM (Jan. 28, 2021), <https://medium.com/what-works-cities-certification/data-driven-government-in-large-small-cities-8d52d5bbaf0>.

⁸⁴ See Wilson & Cong, *supra* note 41, at 6–7 (commenting that “open government data are primarily used by local government staff” and that, if open data is to realize its civic promise, “residents must begin interacting with these resources on a wider basis”).

of groundwater resources.⁸⁵ The legislation attempted to respond to well-documented shortcomings in data about the quality and supply of water in the state.⁸⁶ Before the legislation, the information was incomplete, agencies maintained data in inconsistent ways, and they lacked staffing to turn the data into a format useful for public viewing.⁸⁷ Six years after the law's enactment, more than 1700 data sets about water are now available through the state's centralized open data portal.⁸⁸ Although the platform is publicly accessible, the legislature's primary stated purpose in enacting the Open and Transparent Water Data Act was to facilitate more informed collaboration among government agencies and employees.⁸⁹ There is no question that as public employees have harnessed technology to more mindfully track and report what they do, some efficiencies have resulted.⁹⁰ A 2021 report by Deloitte's Monitor Institute and Bloomberg's What Works Cities initiative concluded after a nationwide series of surveys and interviews that 75 percent of cities are using data to monitor and improve the delivery of services, including making more informed budgeting decisions.⁹¹

Despite the utility of data-sharing to government insiders, the movement has yet to generate widespread public engagement. *Government Technology* magazine examined 55,000 data sets published by 336 different state and local government entities, and found that 90 percent of them had been viewed 2200 times or fewer over the course of their existence.⁹² A 2021 study concluded that "compared with the

⁸⁵ J. Gage Marchini, Note, *Connecting the "Drops" of California Water Data: Chapter 506: The Open and Transparent Water Data Act*, 48 U. PAC. L. REV. 785, 790 (2017).

⁸⁶ See *id.* at 786 (stating that "the greater problem underlying the lack of water is the lack of easily accessible data for California water managers").

⁸⁷ *Id.* at 787, 794–95.

⁸⁸ CA. OPEN DATA PORTAL, <https://data.ca.gov/group/water>.

⁸⁹ Marchini, *supra* note 83, at 790.

⁹⁰ See MONITOR INST. BY DELOITTE / WHAT WORKS CITIES, *Closing the Data Gap: How Cities Are Delivering Better Results for Residents* (July 2021), https://results4america.org/wp-content/uploads/2021/06/Deloitte-WWC-Data-Gap-Report_vFinal-063021.pdf at 3 (hereinafter *Closing Data Gap*) (concluding, after a survey encompassing 44 cities and interviews with 26 city officials, that municipalities are making progress on key open data goals: performance management, public engagement, releasing data, and taking action based on data).

⁹¹ *Id.* at 7–10.

⁹² Andrew Adams, *Open Data Is Here. But What Data Do Governments Publish?*, GOVT. TECH. (Oct. 18, 2022), <https://www.govtech.com/biz/data/open-data-is-here-but-what-data-do-governments-publish>. Interestingly, the article reports the magazine's finding that the two most-viewed data sets out of the 55,000 studied were a log of calls to the Dallas Police Department, and the COVID-19 dashboard of the Connecticut Department of Public Health. *Id.*

managerial benefits within local governments, impacts felt by residents are rare and anecdotal.”⁹³ There is a distinction between being transparent and being accountable. Accountability suggests that someone is actually looking at what the government discloses, providing feedback, and receiving a response.⁹⁴ If the public does not engage with the data, then government information becomes little more than the proverbial tree falling, unobserved, in a civic forest. Some citizens have shown creativity and initiative in tapping government data to build subject-specific websites around pressing needs that the news media and larger public might realistically expect to be looking for, such as in Pittsburgh, where citizens built an app that allows users to assess whether the bridges they use are sound and reliable.⁹⁵ But such success stories are relatively uncommon.⁹⁶

A. Prioritization and Political Will

The police killing of 18-year-old Michael Brown in Ferguson, Missouri, in August 2014 has been called “a galvanizing moment” for the open data movement.⁹⁷ The questionable shooting, and nationwide protests that ensued, put a sharp focus on the dearth of authoritative data about how often police use deadly force and why.⁹⁸ The Obama administration responded by creating the Police Data Initiative, headed by a high-powered group of data scientists

⁹³ Wilson & Cong, *supra* note 41, at 10.

⁹⁴ Tiago Peixoto, *The Uncertain Relationship Between Open Data and Accountability: A Response to Yu and Robinson’s “The New Ambiguity of ‘Open Government’”*, 60 UCLA L. REV. DISC. 200, 203 (2013).

⁹⁵ Amanda Waltz, *Bad Bridges App Helps Commuters Avoid Questionable Pittsburgh Infrastructure*, PITT. CITY PAPER (Feb. 25, 2022), <https://www.pghcitypaper.com/pittsburgh/bad-bridges-app-helps-commuters-avoid-questionable-pittsburgh-infrastructure/Content?oid=21182747>.

⁹⁶ See Erna Ruijter & Albert Meijer, *Open Government Data as an Innovation Process: Lessons From a Living Lab Experiment*, 43(3) PUB. PERFORMANCE & MGT. REV., no. 3, 2020, at 613, 614 (observing that “so far, the impact of open data is rather limited, and this raises questions of why OGD has not lived up to its promise”) (citation omitted).

⁹⁷ Dave Nyczepir, *U.S. Chief Data Scientist Explains Government’s Push for Greater Use of Disaggregated Data*, FEDSCOOP (Sept. 9, 2022), <https://www.fedscoop.com/us-chief-data-scientist-interview/>.

⁹⁸ See Grace E. Leeper, Note, *Conditional Spending and the Need for Data on Lethal Use of Police Force*, 92 N.Y.U. L. REV. 2053, 2057–59 (2017) (discussing need for robust data about lethal force by police in light of Brown case and others like it: “As it stands, we do not know the extent of the issue of police misconduct—or whether backlash against such misconduct might adversely affect police officers’ ability to do their jobs. Without accurate and comprehensive data, we have no way of understanding the scope of the problem, much less identifying and addressing its causes and effects”) (footnote omitted).

and policymakers, whose job was to come up with ways of centrally gathering and displaying critical data sets about the performance of law enforcement functions.⁹⁹ The promise of the Initiative was never realized. Not only did it fail to grow beyond the initial pilot round of police departments, but even many of those early adopters did not actually furnish data for the federal site. For instance, the Initiative claims to have 120 participating law enforcement agencies, but its web page for complaints against police officers—one of the highest-value data sets in contemporary America—contains entries from only six police departments, several of which are empty sets or greatly outdated.¹⁰⁰ The unrealized promise of the Police Data Initiative neatly illustrates the existential challenge that faces the open data movement: The people with the data must be either motivated or forced to share it.

Even if cost and technological know-how is not a prohibitive barrier, government officials may hesitate to share data freely for other reasons including uncertainty of the data's quality and reliability, the belief that hoarding information is an element of their power, or the aspiration to monetize the information.¹⁰¹ On that latter point, copyright law contemplates the possibility that states, counties, or cities might have a protectable interest in intellectual property created by their employees, and at times government entities have earned revenue by selling, for instance, Geographic Information System (GIS) maps or other commercially valuable data.¹⁰² Because of the increasing entanglement between government agencies and private service providers, it is also conceivable that private entities might assert a proprietary interest in data the government possesses.¹⁰³

⁹⁹ Samuel R. Wiseman, *The Criminal Justice Black Box*, 78 OHIO ST. L.J. 349, 356 (2017).

¹⁰⁰ See POLICE DATA INITIATIVE, COMPLAINTS, <https://www.policedatainitiative.org/datasets/complaints/>. For example, the entry for the Philadelphia Police Department links to an archival city web page, <https://www.opendataphilly.org/dataset/philadelphia-police-advisory-commission-complaints>, with complaints covering the period 2009 to 2012.

¹⁰¹ Teresa Scassa, *Public Transit Data Through an Intellectual Property Lens: Lessons About Open Data*, 41 FORDHAM URB. L.J. 1759, 1761 (2014).

¹⁰² See *Pictometry Int'l Corp. v. Freedom of Info. Comm'n*, 59 A.3d 172 (Conn. 2013) (ruling that state can assert copyright over GIS maps and charge FOIA requester licensing fee for making copies); see also Scassa, *supra* note 98, at 1796 (describing how New York's transit agency insisted that an app developer pay \$5,000 plus royalties to license the city's purportedly proprietary transit schedule, and then insisted that the app be taken down when payment was not made).

¹⁰³ See Scassa, *supra* note 98, at 1799 (noting that private vendors may assert intellectual property protection over data about public transit that government agencies generate using the vendors' software or equipment).

Some resistance to gathering and sharing data is undoubtedly politically motivated, as state and local authorities resist being “overseen” by the federal government and seek to avoid publicizing their shortcomings—such as the failure to provide criminal defendants with timely assistance by counsel—that would create pressure for costly reforms.¹⁰⁴ Protection of political turf and entrenched ways of doing business may also play a role. Even a relatively bullish 2021 report on the state of open data acknowledged a continuing challenge in getting different government units within the same geographical area to agree on a shared vision of how data should be gathered and published when they have overlapping responsibility for the same service.¹⁰⁵

B. Privacy Concerns

It is widely accepted that the public can review and copy all manner of government documents containing information that people named in the documents might find embarrassing, including lawsuits, arrest records, and bankruptcy filings. Regardless of the preferences of the individuals involved, the public’s interest in accountability is paramount. For instance, without access to police reports and court records, the public would not know whether police departments are engaging in racial profiling, and journalists would not be able to locate and interview those victimized by profiling.¹⁰⁶ Indeed, some records of great societal value, such as home sales data and property-tax assessments, are universally available even though such records disclose information— “how much did you pay for your house?” – that many would hesitate to share even with close friends. Society tolerates this discomfort because records of property transactions have value both to the marketplace (to facilitate more accurate home pricing) and to journalists and watchdogs (who sometimes draw on home ownership records to document questionable conduct by people in power).¹⁰⁷ Still, the tension between transparency and

¹⁰⁴ Wiseman, *supra* note 96, at 389–90.

¹⁰⁵ *Closing Data Gap*, *supra* note 87, at 15.

¹⁰⁶ See Kate Rabinowitz et al., *How We Calculated the Risks of Walking While Black*, PROPUBLICA (Nov. 16, 2017, 7:59 AM), <https://www.propublica.org/article/how-we-calculated-the-risks-of-walking-while-black> (explaining how journalists used court records of jaywalking citations and other pedestrian offenses to document that Jacksonville sheriff’s deputies were ticketing Black males at greatly disproportionate numbers).

¹⁰⁷ See, e.g., Sally Goldenberg & Joe Anuta, *Burning the Midnight Oil: Eric Adams’ Mysterious Whereabouts Off the Campaign Trail*, POLITICO (June 8, 2021, 5:10 PM), <https://www.politico.com/states/new-york/albany/story/2021/06/08/burning-the-midnight-oil-eric-adams->

personal privacy has been a regular source of conflict, far predating the open data movement. Federal statutes penalize the disclosure of patient information by healthcare providers, student academic records by schools, and driver license data by state motor vehicle agencies.¹⁰⁸ Some degree of informational privacy—the right to control the disclosure of personal confidences without government coercion—is even recognized as being implicit in the Constitution.¹⁰⁹

FOIA statutes typically provide for withholding information that unduly invades personal privacy, such as an exemption in the federal FOIA statute which covers “personnel and medical files and similar files the disclosure of which would constitute a clearly unwarranted invasion of personal privacy”.¹¹⁰ In a landmark 1989 ruling interpreting FOIA, the U.S. Supreme Court concluded that the privacy interest of people named in FBI records outweighed the interests of requesters in seeing FBI compilations of criminal histories, even though each individual piece of data was findable by visiting courthouses and police departments around the country.¹¹¹ States likewise routinely limit disclosure of information on privacy grounds or regulate what users can do with the information they obtain, such as restricting commercial data brokers from reselling records identifying crime victims or people involved in car

mysterious-whereabouts-off-the-campaign-trail-1385412 (drawing on public records to show that leading New York City mayoral candidate provided conflicting information about whether his primary residence was in New York City and which properties he owned); Justin Gray, *Investigation: Marjorie Taylor Greene Filed Homestead Exemptions on 2 Homes, Violating State Law*, WSB-TV ATLANTA (May 14, 2021, 6:56 PM), <https://www.wsbtv.com/news/local/fulton-county/investigation-marjorie-taylor-greene-filed-homestead-exemptions-2-homes-violating-state-law/IXIQMH5PBFBGLCF5ZV44QC6XY/> (using county records to raise questions about whether a controversial member of Congress fraudulently lowered her family’s property taxes).

¹⁰⁸ See 18 U.S.C. §§ 2721–2725 (privacy of driver license records); 20 U.S.C. § 1232g (privacy of education records maintained by federally funded educational institutions); 42 U.S.C. § 1320d(6) (privacy of patient records maintained by healthcare providers or insurers).

¹⁰⁹ See Martin E. Halstuk, *Shielding Private Lives from Prying Eyes: the Escalating Conflict Between Constitutional Privacy and the Accountability Principle of Democracy*, 11 COMM.LAW CONSPECTUS 71, 79–80 (2003) (tracing evolution of Supreme Court’s recognition of constitutionally grounded right to be free from compelled disclosure of intimate personal information); Elbert Lin, *Prioritizing Privacy: a Constitutional Response to the Internet*, 17 BERKELEY TECH. L.J. 1085, 1094–96 (2002) (describing how scholars have struggled to define boundaries of nebulous informational privacy right).

¹¹⁰ 5 U.S.C. § 552(b)(6).

¹¹¹ U.S. Dept. of Just. v. Reporters Comm. for Freedom of Press, 489 U.S. 749 (1989).

accidents.¹¹² As government agencies have widely begun disclosing data online, privacy has been regularly cited as a rationale, or an excuse, for suppressing database entries, even where the data has considerable civic importance.¹¹³

Privacy law is widely misunderstood and over-applied, leading to the withholding of essential information that is not genuinely confidential. For example, it is widely but incorrectly believed that disclosing any information about health or medical care is forbidden by a federal statute, the Health Insurance Portability and Accessibility Act (HIPAA).¹¹⁴ But HIPAA is actually quite narrow.¹¹⁵ Either because of good-faith misinterpretation or bad-faith opportunism, government agencies repeatedly fall back on HIPAA—even at times of public health crisis—to deny the press and public access to data about disease outbreaks and outcomes, even where the possibility of matching the data to identifiable people is far-fetched.¹¹⁶ The more that data is manipulated or

¹¹² Cheryl M. Sheinkopf, Comment, *Balancing Free Speech, Privacy and Open Government: Why Government Should Not Restrict the Truthful Reporting of Public Record Information*, 44 UCLA L. REV. 1567, 1570–71 (1997).

¹¹³ See, e.g., Tessa Weinberg, *Bipartisan Missouri Bills Would Ban Out-of-School Suspensions From Pre-K Through 3rd Grade*, MISSOURI INDEP. (Feb. 5, 2022, 6:00 AM), <https://www.kcur.org/education/2022-02-05/bipartisan-missouri-bills-would-ban-out-of-school-suspensions-from-pre-k-through-3rd-grade> (reporting that Missouri education agency refuses to post demographic information about school suspensions on its open data site, citing privacy concerns, which thwarts analyzing whether suspensions may be levied discriminatorily); Arianna Prothero & Alex Harwin, *Many Online Charter Schools Fail to Graduate Even Half of Their Students on Time*, EDUC. WEEK (Apr. 18, 2019), <https://www.edweek.org/teaching-learning/many-online-charter-schools-fail-to-graduate-even-half-of-their-students-on-time/2019/04> (reporting that federal government masks data about graduation rates at some online charter schools, purportedly on privacy grounds); Kalyn Belsha, *How Illinois Obscures Racial Disparities in School Discipline Data*, CHICAGO REP. (July 18, 2018), <https://www.chicagoreporter.com/how-illinois-obscures-racial-disparities-in-school-discipline-data/> (describing how Chicago schools suppress race and ethnicity information in data about school discipline, leaving advocates to rely on stale data from U.S. Department of Education).

¹¹⁴ 42 U.S.C. § 1320d–6.

¹¹⁵ See Leah R. Fowler, *COVID-19 and the Myth of Health Data Privacy*, 31 KAN. J.L. & PUB. POL'Y 373, 383 (2022) (“The protections HIPAA actually confers are surprisingly limited against the scope of its apparent hold on the American imagination.”).

¹¹⁶ See Al-Amyn Sumar, *Transparency and Access in a Pandemic: Understanding the Impact of HIPAA on Government Disclosures*, 2J. CIVIC INFO., December 2020, at 23, 31 (commenting, in context of government resistance to disclosing data about COVID-19: “The frequent lack of transparency about coronavirus data should be counted among the major failures of our governments’ response to the pandemic. Much of the responsibility for this failure rests with state and local agencies that have, among other things, taken an overly restrictive interpretation of HIPAA’s Privacy Rule.”).

obscured in the interest of privacy, the less useful it becomes.¹¹⁷ To cite one especially troubling example, during the peak of the 2020–21 COVID-19 pandemic, U.S. school authorities habitually refused on privacy grounds to release school-specific data about where outbreaks were occurring—information that would have aided parents in assessing whether their children were at high risk of catching the virus.¹¹⁸

It is undoubtedly true that technology has made it possible to match a mosaic of seemingly unrelated data points to an identifiable individual.¹¹⁹ But the possibility of doing that with government data is typically remote; For instance, the disclosure that two students attending Central High School have COVID-19 would not enable anyone who does not already know the students to reverse-engineer their names. The federal Court of Appeals for the D.C. Circuit made this observation in siding with a journalist who was denied access to data about the prescription-writing practices of military physicians.¹²⁰ Even granting a series of improbable assumptions—such as the assumption that each drug was useful only for a single medical condition—the mere speculative possibility that an outsider might deduce that a particular patient suffered from a particular disease was not enough to justify withholding the data from the journalist.¹²¹ While privacy concerns undoubtedly exist,¹²² the ultimate

¹¹⁷ See Altman et al., *supra* note 6, at 1973–74 (observing that techniques to de-identify records “significantly reduce the utility of data” and “narrows the scope of possible uses of the data”).

¹¹⁸ See Isabella Murray, *State Limits What Iowa Schools Can Share On COVID Info*, IOWA STARTING LINE (Sept. 23, 2020, 8:38 AM), <https://iowastartingline.com/2020/09/23/state-limits-what-iowa-schools-can-share-on-covid-info/> (reporting that state officials told K-12 schools not to share information about COVID cases with anyone other than public health authorities); Shelly Conlon, *State Officials Won’t Say Which Schools Have Coronavirus Cases*, ARGUS LEADER (Sept. 2, 2020, 10:08 AM), <https://www.argusleader.com/story/news/education/2020/09/02/state-officials-wont-say-which-schools-have-coronavirus-cases/5681191002/> (reporting that South Dakota’s health department refused to release public records showing which schools and colleges were reporting positive COVID-19 tests).

¹¹⁹ See Jesse Woo, *Beyond Mosaic Theory: Understanding Privacy Harms in Smart Cities Through a Complexity Theory Lens*, 106 IOWA L. REV. ONLINE 114, 117(2021) (explaining “the central concern of mosaic theory: that connecting a large number of data points will reveal information about an individual even when a single data point would not be considered revealing or privacy-invasive”)(footnote omitted).

¹²⁰ *Arieff v. U.S. Dept. of Navy*, 712 F.2d 1462 (D.C. Cir. 1983).

¹²¹ *Id.* at 1467–68.

¹²² See Borgesius et al., *supra* note 46, at 2090 (stating that “lack of individual control over personal information can lead to subjective and objective privacy harm” and furnishing example of financial donors to

success of open data depends on training government data managers to separate phantom privacy risks from genuine ones and to avoid needless over-compliance that only adds to public skepticism that the government is unduly hidebound.

C. “Data Deserts”: Who’s Counting?

Open data is only as good as the government’s discipline in collecting it. There are persistent and well-documented shortcomings in the way data is gathered, from the highest levels of the federal government down to the smallest local agency. In an especially notable example, the U.S. Department of Justice (DOJ) has long struggled to get states, counties, and cities to cooperate in producing essential criminal justice metrics.¹²³ The most basic measurement of public safety—the number of crimes reported to police—has also been one of the most elusive.¹²⁴ Even where data about public safety is collected and reported, the reports are notoriously incomplete. For example, journalists have regularly reported that, whether accidentally or on purpose, counties are underreporting the number of people who die in jail and how they died.¹²⁵ Whatever details surrounding jail

controversial California ballot initiative who became targets of threats and boycotts after their names were publicized online).

¹²³ See Andrew Ba Tran et al., *As Fatal Police Shootings Increase, More Go Unreported*, WASH. POST (Dec. 6, 2022, 6:30 AM), <https://www.washingtonpost.com/investigations/interactive/2022/fatal-police-shootings-unreported/> (reporting that “at least 2,250” local police or sheriffs’ offices have failed to report data about police shootings to DOJ over the past seven years, leading to federal data that misleadingly indicates fatal police shootings have declined steadily since 2015); Ethan Corey, *DOJ Admits It Has No Idea How Many People Die in Law Enforcement Custody*, APPEAL (Sept. 20, 2022), <https://theappeal.org/doj-deaths-in-custody-failure-missing-deaths/> (citing federal study that indicates, “[s]ince October 2019, the DOJ has missed at least 18 percent of all deaths in state prisons, 39 percent of deaths in local jails, and between 62 and 71 percent of deaths in police custody,” largely attributable to lack of cooperation by states in reporting data).

¹²⁴ See Matthew Yglesias, *What’s Dangerous Is America’s Lack of Crime Data*, BLOOMBERG (Sept. 18, 2022, 5:00 AM), <https://www.bloomberg.com/opinion/articles/2022-09-18/us-crime-rate-is-it-rising-no-one-really-knows> (observing that “incredibly poor” quality of data on incidence of crime makes it impossible to reliably say whether crime is increasing, declining or staying the same).

¹²⁵ See Joshua Vaughn, *Most Deaths in Pa. Jails Went Unreported Despite Rules: ‘It is Appalling’*, PENNLIVE (Feb. 9, 2022, 5:30 AM), <https://www.pennlive.com/news/2022/02/most-deaths-in-pa-jails-went-unreported-despite-rules-it-is-appalling.html> (reporting that, despite federal requirements, Pennsylvania authorities reported only 22 of 56 documented deaths in jail to federal government in 2020); see also Erik Ortiz, *Hundreds of Prison and Jail Deaths Go Uncounted by the Federal Government, Report Finds*, NBC NEWS (Sept. 20, 2022, 5:14 AM), <https://www.nbcnews.com/politics/politics-news/hundreds-prison->

deaths the federal government withholds from the public, the DOJ justifies this as the price of convincing counties to share their records with the agency.¹²⁶ Indeed, data about policing, which is perhaps the most sensitive function that state and local governments perform, can easily seem to be more “hole” than “donut.”¹²⁷

Here again, journalists and citizen watchdogs are scrambling to fill the information vacuum. Researchers have built their own database and map to track deaths in state prisons and county jails across Louisiana, which has one of the nation’s highest per capita incarceration rates.¹²⁸ Reporters for Reuters, the global news wire service, filed more than 1,500 requests for public records to piece together the circumstances of 7,571 deaths in jails that could be documented between 2009 and 2018 in 500 facilities across the country.¹²⁹

The absence of reliable data about the workings of the criminal justice and correctional systems is not because the government lacks the ability to gather data. In fact, Bloomberg commentator Matthew Yglesias points out that the federal government is quite good at gathering granular data when it comes to jobs and the economy, a subject at least as sprawling as crime.¹³⁰ And the federal government *does* effectively gather the data that it chooses to, when gathering is seen as important for solving crimes, including databases about convicted sex offenders or suspected terrorists.¹³¹ What

jail-deaths-go-uncounted-federal-government-report-fin-rcna47954 (recounting findings of congressional investigation, which concluded that, in fiscal year 2021 alone, at least 990 deaths in prisons and jails went uncounted by Justice Department, despite 2014 federal statute, the Death in Custody Reporting Act, that required maintaining a comprehensive national database).

¹²⁶ Peter Eisler et al., *Why 4,998 Died in U.S. Jails Without Getting Their Day in Court*, REUTERS (Oct. 16, 2020, 11:00 AM), <https://www.reuters.com/investigates/special-report/usa-jails-deaths/>.

¹²⁷ See Wiseman, *supra* note 96, at 350 (“We do not know exactly how many people in the United States are in jail or prison, what they were charged with, the charges for which they were convicted, how long they awaited trial in jail, and when and whether they had an attorney, among other basic data points. Yet we live in the era of so-called ‘big data.’”).

¹²⁸ INCARCERATION TRANSPARENCY, <https://www.incarcerationtransparency.org/> (last visited Jan. 22, 2023).

¹²⁹ Eisler et al., *supra* note 122.

¹³⁰ Yglesias, *supra* note 120. See also Grace C. Leeper, *Conditional Spending and the Need for Data on Lethal Use of Police Force*, 92 N.Y.U. L. REV. 2053, 2054-55 (2017) (enumerating examples of granular data amassed by the federal government in areas other than policing, but observing that data is absent or incomplete as to how often police use deadly force, how often they stop and frisk citizens, or even how many police departments exist in the United States).

¹³¹ Wiseman, *supra* note 96, at 395.

is missing in the criminal justice field is motivation. State and local governments, where the data originates, have little incentive to cooperate with federal agencies, and federal bureaucrats suffer no consequences for data deserts that they can blame on state and local obstructionism. Very few data-collection programs come with a meaningful “carrot” and “stick” to incentivize compliance and penalize noncompliance. For example, the U.S. Department of Education had evidence for over a decade that colleges and universities were systematically understating the frequency of serious crimes in federally mandated annual reports; yet almost none of the institutions were ever penalized, unless journalists or outside watchdog groups exposed an especially egregious falsehood.¹³² Without the force of legal compulsion, there is rarely any self-interest for government officials to publish data voluntarily, especially because it may expose mismanagement or deteriorating quality of life, and it provides a baseline from which the press and public can hold officials accountable, such as by documenting that crime is going up or that student test scores are going down.

Open data is useful only if it is reliable. Opening and publishing known incomplete data is arguably worse than publishing none at all. It may lull the public into underestimating the severity of a problem, such as whether the local jail is treating detainees inhumanely. As long as the government does not prioritize tracking its activities in a way that lends itself to meaningful, apples-to-apples comparisons, open data portals run the real risk of a “garbage in, garbage out” problem.

D. Lack of Audience Engagement

For government-disclosed information to be of value, it must actually reach the intended audience and fulfill the public’s perceived information needs.¹³³ Additionally, because most citizens are not trained data scientists, open data is valuable only if it is posted in a format that is both easy to use and easy to comprehend.¹³⁴ Critics have suggested that the first generation of open data platforms is largely accounting-

¹³² See Tyler Kingkade, *17 Colleges Fell Short on Campus Safety, But the Education Department Didn’t Tell the Schools*, BUZZFEED NEWS (Apr. 5, 2018, 9:22 AM), <https://www.buzzfeednews.com/article/tylerkingkade/clery-act-campus-safety-shortcomings-betsy-devos>; see also *Campus Insecurity: Inside the Investigation*, COLUMBUS DISPATCH (Nov. 23, 2014, 2:30 PM), <https://www.dispatch.com/story/news/crime/2014/11/24/campus-insecurity-inside-investigation/23674233007/>.

¹³³ Peixoto, *supra* note 91, at 204.

¹³⁴ *Id.*

driven, an inventory of the information that states and cities hold, rather than user-experience-driven, organized around the way people use data and their priorities.¹³⁵

In a study of the experience with three open data portals in New England states, researchers Bonnie Paige and Lianne Freund identified the concern that unless portals are designed with user value in mind they could become “empty scaffolding” that satisfies statutory mandates or political imperatives without actually providing value to users.¹³⁶ It was observed that potential users may be deterred from taking time to visit the sites both because of actual lack of expertise and perceptions or misperceptions that the task will be difficult.¹³⁷ In inspecting portals from Maine, New Hampshire and Vermont, Paige and Freund found neither consistency among states in which data sets they featured, nor any public-facing explanation about why certain data sets were selected for inclusion.¹³⁸ They concluded that “user action has been minimal” in each of the states, explaining: “It is difficult to discover these portals via other government websites due to inadequate linking and they have no social media presence, so most traffic would have to come from search engine users seeking open data.”¹³⁹

One critical choice for data custodians is the type of data to collect and display. Cabitza, Locor, and Batini recommend prioritizing open data along a continuum of three use-case factors: (1) criticality, for instance, the need to obtain hospital care; (2) connection, or the degree to which users feel that data is important in their lives or those of close loved ones; and (3) urgency, or the usefulness of the data in making important decisions, such as deciding which hospital to go to for a stroke.¹⁴⁰

Another critical choice is at what stage to involve the end-users of data in the design process. The nonprofit Sunlight Foundation, a leading authority on open government data until its closure in 2020,¹⁴¹ issued a report in

¹³⁵ Federico Cabitza et al., *Making Open Data More Personal Through a Social Value Perspective: A Methodological Approach*, 22 INFO. SYS. FRONTIERS 131, 131 (2020). See also Wilson & Cong, *supra* note 41, at 10 (commenting that what governments choose to make available online is largely “supply-driven” based on what is available today, rather than considering “the needs of priorities of non-government users of those data...”).

¹³⁶ Bonnie Paige & Lianne Freund, *A Case Study of New England Open Data Portals*, 11(1) JEDEM 59, 60 (2019).

¹³⁷ *Id.* at 62.

¹³⁸ *Id.* at 65-66.

¹³⁹ *Id.* at 73.

¹⁴⁰ Cabitza et al., *supra* note 131, at 137.

¹⁴¹ Ruth McCambridge, *Becoming an Ancestor: Sunlight Foundation Closes Its Own Shutters*, NONPROFIT Q. (Sept. 30, 2020),

2017 encouraging municipalities to adopt a tactical engagement strategy to promote greater public use of data.¹⁴² The process starts with observing community needs to identify a subject matter to focus on, then interviewing stakeholders to develop a use case for data, coordinating with community partners to identify target users, and then working with those partners to roll out the data along with interpretive guides.¹⁴³ Sunlight’s recommended process contemplates opening data as an ongoing and organic process, with feedback from civic hackers and other knowledgeable users incorporated into improving the functionality of the sites, and sharing of successful data stories from community members who have put the information to use.¹⁴⁴

IV. RECOMMENDATIONS

America’s experience with data dashboards during the height of the COVID-19 pandemic is a microcosm of what happens when the public is hungry for real-time data. Technology makes it possible to deliver the data, but the results do not always answer the public’s most pressing questions in a useful way.

As the severity of COVID-19 became grimly apparent during the spring and summer of 2020, the public and press clamored for updates, looking for answers to questions such as: How many people are testing positive, how many are being hospitalized, how many are dying, are there hotspots in congregate living facilities such as nursing homes, prisons, or college dorms? But when government employees were largely sent home to work remotely, fulfillment of FOIA requests slowed down or sometimes entirely halted.¹⁴⁵ Some states took advantage of state-of-emergency declarations to suspend the statutory deadlines in their open-government laws, giving them essentially an indefinite period to delay compliance without consequences.¹⁴⁶ Critics noted the irony that data was least accessible when it was most urgently needed: “A public health crisis like the COVID-19 pandemic

<https://nonprofitquarterly.org/becoming-an-ancestor-sunlight-foundation-closes-its-own-shutters/>.

¹⁴² *Tactical Data Engagement*, SUNLIGHT FOUND. (Sept. 2017), <https://communities.sunlightfoundation.com/methodology/>.

¹⁴³ *Id.*

¹⁴⁴ *Id.*

¹⁴⁵ See Adam A. Marshall & Gunita Singh, *Access to Public Records and the Role of the News Media in Providing Information About COVID-19*, 11 J. NAT’L SEC. L. & POL’Y 199, 202 (2020) (reporting that the FBI and State Department were among the agencies that suspended processing of FOIA requests during the worst of the pandemic).

¹⁴⁶ *Id.* at 204.

warrants more transparency—not less—to ensure that the public has access to information about the nature of the disease sweeping our nation and the government’s response to it.”¹⁴⁷

Unable to dislodge government records promptly, the press and public became reliant on state-created data dashboards, which displayed statistics compiled from public-health authorities and healthcare providers.¹⁴⁸ But the data did not always tell a complete and accurate story, and sometimes the inaccuracy was by design. In New York, Governor Andrew Cuomo’s administration was accused of protecting the nursing home industry by distorting how many people had died in the facilities, understating the death toll by claiming thousands of patients had actually died after being taken to the hospital.¹⁴⁹ In Florida, Governor Ron DeSantis’ administration discontinued providing daily details about COVID-19 during the summer of 2021, even though the disease was continuing to ravage the country, resulting in a 105-day outage during which Florida fatalities went undisclosed.¹⁵⁰ Nationwide, authorities falsely cited HIPAA rules about medical privacy when withholding information about trends in infections and deaths within schools, hospitals, and nursing homes, as if a person who did not already know the patients’ identities could magically deduce them through numerical tabulations alone.¹⁵¹

As so often happens, news reporters stepped in to fill the void. A team of journalists at *The Atlantic* magazine began combing through county-level reports from public health authorities to build a regularly updated tracker that drew on expertise from data scientists to furnish reporting that was

¹⁴⁷ *Id.* at 206.

¹⁴⁸ See generally Sara Kizer, *Visualizing a Way Forward: Lessons Learned and Successful Strategies from COVID-19 Data Dashboard Implementation*, 61 JURIMETRICS J. 469, 475-76 (2021) (commenting on the proliferation of federal and state dashboards to track COVID-19 and identifying recurring quality-control challenges with public health data).

¹⁴⁹ J. David Goodman & Danny Hakim, *Cuomo Aides Rewrote Nursing Home Report to Hide Higher Death Toll*, N.Y. TIMES (Sept. 23, 2021), <https://www.nytimes.com/2021/03/04/nyregion/cuomo-nursing-home-deaths.html>.

¹⁵⁰ Ian Hodgson, *For 105 Days, COVID’s Death Toll in Florida Counties Went Missing*, TAMPA BAY TIMES (Oct. 17, 2021), <https://www.tampabay.com/news/health/2021/10/17/for-105-days-covids-death-toll-in-florida-counties-went-missing/>.

¹⁵¹ See Sumar, *supra* note 112, at 27 (noting a “striking” inconsistency in the application of federal privacy law among states—and even between counties in the same state—to deny access even to data about COVID cases devoid of any personal identifiers).

more thorough than that offered by the federal government.¹⁵² *The New York Times* ambitiously built a county-by-county tracker of every known COVID-19 case, hosted on the open-source platform Github, which was meant to be more granular than the federal Centers for Disease Control’s official state-by-state count.¹⁵³ At Columbia University, journalism students and faculty launched “Documenting COVID-19,” a repository of publicly available data and documents that Columbia researchers accumulated through nationwide waves of FOIA requests and made available for community journalists to adapt and localize.¹⁵⁴ The Columbia project birthed data-driven news stories at outlets across the country, demonstrating the usefulness of well-curated data sets in powering local accountability journalism.¹⁵⁵

If government agencies do not have high-quality data about their operations to share with the public, that absence may be the proverbial canary in the coal mine indicating deeper management problems. An agency that does not readily know the volume and type of work it is handling, in which direction trends are heading, or how the agency’s work measures up against comparable agencies elsewhere, is in no position to assess its own effectiveness. The very absence of data—such as a failure to track trends surrounding police use of force—can be a red flag indicating that the agency’s management is failing to ask important questions.

Over the years, scholars and advocates have attempted to distill best practices for gathering and producing government data in a way that the public will find useful and engaging. In 2007, O’Reilly Media and Public.Resource.org convened a working group that produced eight guiding principles for open data: It must be complete, primary, timely,

¹⁵² Emily Sohn, *How the COVID Tracking Project Fills the Public Health Data Gap*, COLUM. JOURNALISM REV. (Mar. 24, 2020), https://www.cjr.org/the_profile/covid-tracking-project.php.

¹⁵³ *We’re Sharing Coronavirus Case Data for Every U.S. County*, N.Y. TIMES, <https://www.nytimes.com/article/coronavirus-county-data-us.html> (last visited Jan. 21, 2023).

¹⁵⁴ *The Launch of the ‘Documenting COVID-19’ Public Records Repository*, COLUMBIA UNIV. BROWN INST. FOR MEDIA INNOVATION (July 13, 2020), <https://brown.columbia.edu/documenting-launch/>.

¹⁵⁵ See, e.g., Mary Hynes, *COVID Case Cluster Hits Vaccinated Las Vegas Hospital Workers*, LAS VEGAS REVIEW-JOURNAL (July 13, 2021, 3:38 PM) <https://www.reviewjournal.com/life/health/covid-case-cluster-hits-vaccinated-las-vegas-hospital-workers-2398382/> (using Columbia data to examine breakthrough cluster of infections among vaccinated hospital employees); Jackie Botts et al., *Why Few Farmworkers Isolate in California’s Free COVID-19 Hotel Rooms*, CAL MATTERS (June 15, 2021), <https://calmatters.org/california-divide/2020/12/why-few-farmworkers-isolate-in-californias-free-covid-19-hotel-rooms/> (citing Columbia data in questioning effectiveness of state program offering farmworkers free hotel rooms for quarantine, which few have used).

accessible, machine-readable, available in a nondiscriminatory way, non-proprietary, and license-free.¹⁵⁶

One of the challenges facing the open data movement is to account for the varying types of potential users and their needs. For instance, the average commuter who uses the same bus route every day to get back-and-forth between home and work probably has no use for comparative data showing ridership of every route in the county, but that data might be of great use to scholars, urban planners, or even commercial marketers.¹⁵⁷ The highest-value data sets enable the public to evaluate the way the government is, or is not, delivering essential services and performing essential duties.¹⁵⁸ This would counsel in favor of prioritizing, for instance, a database showing how rapidly emergency medical crews respond to 911 calls ahead of a database of dogs registered with municipal licensing authorities.¹⁵⁹ If open data portals are not visually prioritizing data sets of highest civic value, only the most determined power-searchers will locate the data on their own.

Affirmative steps can be taken both to improve the quality of open data portals and to better prepare journalists, as well as other citizens, to make use of them. On the user side, college journalism programs already are more mindfully investing in the discipline of data journalism, recognizing that this subspecialty is where graduates are most likely to find gainful employment in a difficult market.¹⁶⁰ Along with universities, public libraries are a promising space in which trainees of all ages might obtain assistance in locating and analyzing data from their states and communities.¹⁶¹ More

¹⁵⁶ Whittington et al., *supra* note 28, at 1925-26.

¹⁵⁷ Scassa, *supra* note 98, at 1768. *See also* Cabitza et al., *supra* note 131, at 131-33 (suggesting that open data should not merely entail posting the results of predetermined queries to existing databases, but rather should be designed around the social utility of the data to different categories of users).

¹⁵⁸ *See* Borgesius, *supra* note 46, at 2127 (“The more relevant data are to key aspects of democratic participation, the stronger the case for release as open data.”).

¹⁵⁹ *See* *Bozzi v. City of Jersey City*, 258 A.3d 1048 (N.J. 2021) (holding that database of names and addresses of dog-license holders was disclosable under New Jersey Open Public Records Act and not subject to withholding on personal privacy grounds).

¹⁶⁰ *See, e.g.*, Norm Lewis, *How Data Journalism Changed the University of Florida*, MEDIASHIFT (June 7, 2016), <https://mediashift.org/2016/06/how-data-journalism-changed-the-university-of-florida/> (describing how University of Florida journalism program went from zero data journalism course offerings to five in a three-year period by reassigning existing faculty to teach coding skills).

¹⁶¹ Pittsburgh, recognized as an early adopter in the open data field, has already piloted “data literacy” training for local citizens through its library system. *See* Elizabeth Monk, *Fulfilling the Promise of Open Data through Data*

widely available and affordable training could help bridge what some researchers have called the “digital divide” that inhibits less tech-savvy citizens from making use of government data.¹⁶²

To address common deficiencies afflicting open data portals, no shortage of expert recommendations exists. The nonprofit organization U.S. Open Data conducted an assessment ending in 2016 analyzing the condition of open data websites in all 50 states and Puerto Rico.¹⁶³ The organization’s report concluded that the best-functioning portals came from agencies that employed chief data officers dedicated to gathering data sets.¹⁶⁴ The report also recommended avoiding “data territorialism” by promoting cross-linking among agencies, thus increasing the odds that a searcher will stumble upon useful information.¹⁶⁵

One clear takeaway from the first decade’s experience with open data is that simply throwing existing data sets onto a website without curation, prioritization, or quality control is a strategy for failure. Only the most motivated super-researcher is likely to dig deeply through a collection of hundreds of unrelated data sets of varying quality, importance, and completeness. Drawing on the past decade’s worth of research and experience, the authors recommend a thought exercise for managers of government data to adopt when prioritizing which data sets they decide to feature. First, managers should consider notions of urgency: Does the data address itself to a pressing community need or problem? Are health and safety concerns implicated, or is the data simply a curious trivia point? Second, managers should consider notions of actionability: Does having the data equip citizens to better take advantage of government services? Can members of the public usefully apply the data to discharge their civic oversight role by evaluating how the government is performing?¹⁶⁶ The third consideration is verifiability: Is the

Literacy Training, MICROSOFT (Aug. 1, 2017), <https://blogs.microsoft.com/newyork/2017/08/01/fulfilling-the-promise-of-open-data-through-data-literacy-training/>.

¹⁶² See Wilson & Cong, *supra* note 41, at 7 (identifying both access to specialized software and training as barriers to widespread public adoption of open data).

¹⁶³ Sarah Schacht, *Open Data Census Provides a To-Do List for States*, GOV’T TECH. (Mar. 9, 2016), <https://www.govtech.com/data/open-data-census-to-do-list-for-states.html>.

¹⁶⁴ *Id.*

¹⁶⁵ *Id.*

¹⁶⁶ See Cabitza *et al.*, *supra* note 131, at 133 (describing “social value” of open data by reference to “its availability for and usability by” the community). The authors regard data as high-value when “users can exploit data to achieve some relevant goal of theirs” and apply it to create value, whether civic or aesthetic. *Id.* at 132-33. The Obama administration, in a 2009

information amendable to third-party validation, or are the underlying source documents from which the data was derived inaccessible?¹⁶⁷

Prioritizing data sets that are verifiable is very important for the data's usefulness to reporters and the public. It means choosing data sets that are not easily subject to untraceable human manipulation. For example, traffic tickets generated by automated red light cameras or speed cameras are data points that law enforcement agencies cannot easily fake without detection (if, for no other reason, than the paper trail that will exist at municipal courts as the tickets are paid or adjudicated). Other data sets, such as a health department's statistics about how many vaccine doses are administered or a library system's statistics about how many books are returned belatedly, have inherent reliability because there is no obvious self-interested motive to manipulate them.

To overcome the privacy paralysis previously discussed that inhibits government agencies from sharing data, the starting point should be data that reveals something about the *government*, as opposed to data that reveals only information about *individuals*. The U.S. Supreme Court has defined oversight of government as the foundational purpose of FOIA laws.¹⁶⁸ Reasonable people can dispute whether data about sexually transmitted diseases in a small population is personally identifiable in needlessly embarrassing ways, but data about how often police use force, or how they are (or are not) disciplined, is inarguably a matter that reveals information about the government and is of public concern and therefore not legitimately withheld on personal privacy grounds.¹⁶⁹

directive to federal agencies, used a definition of "high-value" data sets that included those which add to the public's understanding of government or "create economic opportunity." Yu & Robinson, *supra* note 47, at 198.

¹⁶⁷ See Ryan Little, *Monsters: Fancy Government Charts With No Raw Data*, MOTHER JONES (Dec. 27, 2021), <https://www.motherjones.com/politics/2021/12/monsters-fancy-government-charts-with-no-raw-data/> (decrying government agencies' practices of publishing selective summaries of data without making underlying records available for verification, by using example of visuals posted by U.S. Customs and Border Protection that purport to show how many migrants are being stopped at the U.S.-Mexico border, without furnishing supporting documentation).

¹⁶⁸ See U.S. Dep't of Just. v. Reporters Comm. for Freedom of Press, 489 U.S. 749, 773 (1989) ("Official information that sheds light on an agency's performance of its statutory duties falls squarely within [FOIA's] statutory purpose.").

¹⁶⁹ See Barry Friedman & Maria Ponomarenko, *Democratic Policing*, 90 N.Y.U. L. REV. 1827, 1849 (2015) ("Policing agencies fail to collect – or

If there is concern that records will be unduly invasive of privacy if produced, the time to think about that concern is at the document or database's inception. For example, some state statutes already contemplate that the public is entitled to see the front page of multipage police reports, which contains the essential who-what-where-when information about crimes, while details about the investigation contained in latter pages of the reports can be kept confidential.¹⁷⁰ There is no reason government recordkeeping cannot track this front page approach more generally, so that government agencies create records from their inception with an eye toward ease of public production.¹⁷¹ For example, if a municipality keeps a log of complainants who report violations to the Code Enforcement Office, the name and contact information for the caller could be segregated into a redactable cell on a spreadsheet, separate from the substantive information about the nature of the complaint and its outcome, which are the data points that the public cares about.

Finally, the inherent conflict-of-interest at the heart of open data must be acknowledged: the possibility that government agencies will hesitate to candidly disclose data that reflects poorly on their own performance. Perhaps for this reason, the first generation of open data portals has been built largely around the imperative of enhancing government efficiency, as opposed to the freedom-of-information movement, which has always emphasized the right to know as a way to check government abuses.¹⁷² One way to deal with this inherent limitation on the trustworthiness of self-reported data is to subject data to auditing in the same way

publicize if they do – even basic data that would allow for public supervision and sound cost-benefit analysis.”).

¹⁷⁰ See CAL. GOV'T CODE § 6254(f)(2)(A) (specifying that the public is entitled to a subset of information contained in police reports: “[T]he time, substance, and location of all complaints or requests for assistance received by the agency and the time and nature of the response thereto, including, to the extent the information regarding crimes alleged or committed or any other incident investigated is recorded, the time, date, and location of occurrence, the time and date of the report, the name and age of the victim, the factual circumstances surrounding the crime or incident, and a general description of any injuries, property, or weapons involved.”).

¹⁷¹ See Marijin Janssen et al., *Transparency-By-Design as a Foundation for Open Government*, 11(1) TRANSFORMING GOV'T: PEOPLE, PROCESS & POL'Y 2, 7 (2017), <https://doi.org/10.1108/TG-02-2017-0015> (“Instead of adding a layer for the opening of data, it would be better to organize the data in such a way that it can be disclosed in a more useful way and easily opened without needing additional cumbersome steps.”).

¹⁷² See Berliner et al., *supra* note 77, at 876 (asserting that the arrival of the open data movement “has marked a *relative* shift towards values of efficiency, compared with the more focused FOI emphasis on democratic values that preceded it.”).

that government financial statements are audited. Audit reports expose misstatements in government financial records, whether accidental or purposeful. The existence of an independent review of financial transactions can both deter wrongdoers and reassure the public that abuses will be caught. Auditing agencies already are broadening their attention beyond just financial transactions and also reviewing the effectiveness of government cybersecurity practices as well.¹⁷³ It would be relatively simple for auditors to try to trace the origin of a sample of published datapoints, in the same way that they already sample purchases and contracts, to make sure that the claims made by government agencies check out to be truthful.

V. CONCLUSION

In May 2022, Streetsblog NYC, an online news outlet affiliated with a nonprofit organization that promotes alternative transportation, published the results of a massive analysis of nearly 1 million records of auto accidents over a six-year period, gathered from the NYC Open Data portal.¹⁷⁴ Reporter Jesse Coburn's analysis concluded that intersections near schools, especially schools serving large populations of nonwhite children, were disproportionately prone to car collisions with pedestrians.¹⁷⁵ The article linked to the NYC Open Data site so inquisitive readers could examine the underlying data for themselves, but Coburn did not just regurgitate raw data. She used it as the starting point for reporting that brought in the voices of transportation experts, community leaders, and the affected school children themselves, one of whom told Streetsblog: "I had a few cars try to run me over. You just [get] used to it."¹⁷⁶ Public-service

¹⁷³ See *Secretary of State's Cybersecurity Audit Finds Gaps in Oregon Department of Corrections Defenses*, KTVZ (July 20, 2022, 10:17 AM), <https://ktvz.com/news/government-politics/2022/07/20/secretary-of-states-cybersecurity-audit-finds-gaps-in-oregon-department-of-corrections-defenses/> (reporting that the Oregon Secretary of State conducted a series of eight agency specific audits examining vulnerability to cyberattacks); see also Nick Cahill, *Audit Rips California's Cybersecurity Watchdog over Lack of Oversight*, COURTHOUSE NEWS SERV. (Jan. 18, 2022), <https://www.courthousenews.com/audit-rips-californias-cybersecurity-watchdog-over-lack-of-oversight/> (describing findings of California's state audit agency, which concluded that "California remains woefully behind on ensuring state agencies are protecting critical public information from hackers").

¹⁷⁴ Jesse Coburn, *'Always Scared': Dangerous Streets Outside City Schools Threaten Children*, STREETS BLOG NYC (May 24, 2022), <https://nyc.streetsblog.org/2022/05/24/danger-zones-chaotic-school-streets-threaten-city-children/>.

¹⁷⁵ *Id.*

¹⁷⁶ *Id.*

reporting of this kind demonstrates what is possible when two predicate conditions exist: A robust open data portal with regularly updated information about issues the public cares about, and well-trained reporters with the time and expertise to use computer-assisted reporting to interpret large data sets and make them relevant to problem solving. New York City is fortunate to have both, but few communities are as well-served.

While journalists and government officials are often seen as occupying adversarial positions, they share a common interest in rebuilding the public's frayed trust in all institutions, government and media alike.¹⁷⁷ One method of establishing trust with skeptical audiences is transparency: Rather than simply asserting that something is true, government officials can, and should, take advantage of technology to furnish the factual basis, if any exists, from which their assertions are derived.¹⁷⁸ If Americans are not basing their opinions on a shared set of verifiable facts, demagogues and disinformation purveyors will prosper—as perhaps best illustrated by the pervasive (and widely accepted) political trope that violent crime is soaring, all statistical evidence pointing to the contrary.¹⁷⁹

America is not going back to a time when households widely subscribed to daily newspapers and made the six o'clock television newscast a family ritual. People expect to get information in real time, free of charge, and with easy online access through their mobile devices. Open data portals have a real potential to meet the need for a baseline of factual information on which activists, journalists, and policymakers can mutually rely. But, for the information to have value it must be verifiable and consistent.

As mainstream news organizations have diminished in reach and influence, various types of replacements have

¹⁷⁷ See Megan Brenan, *Americans' Trust in Government Remains Low*, GALLUP (Sept. 30, 2021), <https://news.gallup.com/poll/355124/americans-trust-government-remains-low.aspx> (describing results of nationwide survey of 1,005 adults during September 2021 that found only 39% expressed trust in the federal government to solve problems, and that confidence in all branches of the U.S. government, at all levels, has steadily declined since the 1970s).

¹⁷⁸ See Rajav Desai, *3 Ways Governments Can Regain Citizens' Trust*, ROUTE FIFTY (June 13, 2022), <https://www.route-fifty.com/management/2022/06/governments-regain-citizens-trust/368116/> ("When governments are transparent, it's easier to hold them accountable. And when they know they can be held accountable, they tend to operate more efficiently.").

¹⁷⁹ Maggie Koerth & Amelia Thomson-DeVeaux, *Many Americans Are Convinced Crime Is Rising in The U.S. They're Wrong.*, FIVETHIRTYEIGHT (Aug. 3, 2020, 6:00 AM), <https://fivethirtyeight.com/features/many-americans-are-convinced-crime-is-rising-in-the-u-s-theyre-wrong/>.

developed to fill people’s craving to share and consume information. Discussion forums such as Reddit, and the localized bulletin-board app NextDoor, give nonprofessionals the ability to share their observations with minimal centralized editing or curation, for better or worse.¹⁸⁰ With amateurs now flooding social media and the blogosphere with information of varying quality and believability, the accessibility of reliable government data, assuming it is indeed reliable, can be an asset in the fight against both disinformation and misinformation.

Open government data can never, of course, entirely replace entrepreneurial data-gathering by journalists.¹⁸¹ Government agencies predictably will not volunteer data that discloses misconduct by public officials. Nor will agencies always choose to prioritize opening the data that journalists’ independent news judgment tells them the public needs. Over-reliance on government-volunteered data may even work at cross-purposes with the goal of data journalism to engender greater public trust.¹⁸² For these reasons, well-functioning open data portals should be part—but only part—of a comprehensive strategy to maximize the accessibility of high-quality information, which also necessarily includes expediting compliance with FOIA laws and minimizing reliance on FOIA exemptions.

There is a considerable amount of reasonable skepticism about the proper role of government in directly subsidizing struggling news outlets. Skeptics decry the obvious risk that subsidies could be selectively awarded to media that shares the bias of the decisionmakers in powers, or could be withheld in retaliation for unflattering coverage.¹⁸³ Still, historians note that indirect subsidies have existed throughout U.S. history, such as tax breaks and preferred mail delivery rates.¹⁸⁴ Investing in reliable and transparent open

¹⁸⁰ See Edward Ongweso Jr., *Nextdoor Knows Its Users Are Racist*, VICE (Apr. 20, 2021, 9:05 AM), <https://www.vice.com/en/article/pkbymk/nextdoor-knows-its-users-are-racist> (critiquing Nextdoor as “[a] social network that combines the worst parts of a neighborhood watch and a cliquey message board.”).

¹⁸¹ See Noveck, *supra* note 5, at 32 (commenting that “open data cannot be the sole tool for sparking government information-sharing” and that it operates as a complement to FOIA).

¹⁸² Zamith, *supra* note 20, at 472.

¹⁸³ See Jeff Jacoby, *A Free Press Doesn’t Take Government Handouts*, BOSTON GLOBE (Oct. 24, 2021, 10:00 AM), <https://www.bostonglobe.com/2021/10/24/opinion/free-press-doesnt-take-government-handouts/> (“If keeping the subsidy spigot open means staying on good terms with certain politicians, how likely will the paper be to cover stories the politicians don’t want covered?”).

¹⁸⁴ See Joshua P. Darr, *Government Subsidies to Save Local News*, NIEMAN LAB (Dec. 21, 2021), <https://www.niemanlab.org/2021/12/government-subsidies-to-save-local-news/> (“The government has always subsidized

data production could be viewed as an indirect way of supporting the work of local journalists without the risk of viewpoint-based retaliation; news organizations would save both on labor costs in gathering the data and on fees assessed for fulfilling FOIA requests.¹⁸⁵

The First Amendment is recognized as contemplating active public participation in the process of formulating government policy. But when the government hoards information, it is harder for citizens to have informed and effective participation.¹⁸⁶ Opening data levels the knowledge gap between the regulators and the regulated, so each participant can enter the policymaking process equally well-prepared and drawing from a shared set of facts. Of course, tables of statistics by themselves will not fully empower the press and public to keep watch over the delivery of government services. As the founders of the federal Police Data Initiative have pointed out, disclosing data must be part of a comprehensive transparency strategy that also includes giving the public access to, among other things, internal agency rulebooks, records of agency adjudications, and other documents that fully illuminate the story behind the numbers.¹⁸⁷

Although traditional journalism as a full-time salaried profession is eroding, there will always be a need for trusted intermediaries to validate the information that government agencies push out, and to pry loose the information that agencies withhold.¹⁸⁸ In one especially telling anecdote, the

local news. ...Politicians and government agencies regularly grant the media tax breaks and exempt them from some categories of labor laws and interstate commerce restrictions.”).

¹⁸⁵ *But see* Zamith, *supra* note 20, at 477 (cautioning that “the proliferation of data journalism in particular may make journalism even more vulnerable to a dependence on a new type of information subsidy – data – that can be exploited due to news organizations’ inability to collect their own large datasets.”).

¹⁸⁶ *See* *Martin v. City of Struthers, Ohio*, 319 U.S. 141, 143 (1943) (“The authors of the First Amendment knew that novel and unconventional ideas might disturb the complacent, but they chose to encourage a freedom which they believed essential if vigorous enlightenment was ever to triumph over slothful ignorance. This freedom embraces the right to distribute literature ... and necessarily protects the right to receive it.” (citation omitted)).

¹⁸⁷ Clarence Wardell & Denice Ross, *Police Data Belongs to the People*, BOSTON GLOBE (June 12, 2020, 4:00 AM), <https://www.bostonglobe.com/2020/06/12/opinion/police-data-belongs-people/>.

¹⁸⁸ *See* Andrew Young et al., *A Layered Approach to Documenting How the Third Wave of Open Data Can Provide Societal Value*, OPEN DATA POL’Y LAB (Aug. 25, 2021), <https://opendatapolicylab.org/articles/the-onion-model-a-layered-approach-to-documenting-how-the-third-wave-of-open-data-can-provide-societal-value/>; *see also* Peixoto, *supra* note 91, at 204-05 (observing that “most citizens depend on technically skilled and

former chief data officer for the mayor of Los Angeles related that the best source of local crime data is not the city government but the *Los Angeles Times* newspaper, because the city's validation process is so slow, contributing to a lag of as much as two years.¹⁸⁹ Tomorrow's journalists may well have open data auditor in the top line of their job descriptions. Without independent validation, more information is just more information; but with independent validation, more is better.

resourced individuals to mediate their access to adaptable data" and commenting on the importance of a "vigorous free press" in processing civically essential data).

¹⁸⁹ Abhi Nemani, *Data & Dashboards: The Linchpin of the Smart City*, 85 UMKCL. REV. 973, 977 (2017).