

# The Criminal Justice Black Box

SAMUEL R. WISEMAN\*

*“Big data”—the collection and statistical analysis of numerous digital data points—has transformed the commercial and policy realms, changing firms’ understanding of consumer behavior and improving problems ranging from traffic congestion to drug interactions. In the criminal justice field, police now use data from widely dispersed monitoring equipment, crime databases, and statistical analysis to predict where and when crimes will occur, and police body cameras have the potential to both provide key evidence and reduce misconduct. But in many jurisdictions, digital access to basic criminal court records remains surprisingly limited, and, in contrast to the civil context, no lucrative market for the data (apart from that for background checks) exists to induce the private sector to step in to fill the gap. As a result, bulk criminal justice data is largely limited to survey data collected by the Bureau of Justice Statistics. Unlocking the “black box” by uniformly collecting and reporting basic, anonymized data from criminal cases—including, e.g., the charges, pretrial release decision, appointment of counsel, and case disposition—would have significant benefits. It would allow researchers, reformers, and government actors to both more effectively study the system as a whole and to more easily identify jurisdictions violating the Constitution by, for example, routinely denying counsel or pretrial release and imprisoning defendants for inability (rather than unwillingness) to pay a fine or fee. This Article documents this problem, explores its causes, and proposes a solution, arguing that the federal government should form a framework for the uniform collection of anonymized local, state, and federal criminal justice data. While participation in this uniform system is likely to be incremental, even partial data would improve understanding of the system as a whole and aid efforts to enforce well-established, but frequently violated, constitutional rights.*

## TABLE OF CONTENTS

I.	INTRODUCTION .....	350
II.	THE USES OF BULK CRIMINAL JUSTICE DATA .....	359
	A. <i>Data and Systemic Problems</i> .....	359
	B. <i>Data and Local Problems</i> .....	363
	1. <i>Bail and Pretrial Detention</i> .....	364

---

\* McConnaughay and Rissman Professor, Florida State University College of Law. B.A., Yale University; J.D., Yale Law School. Many thanks to Wayne Logan, Avlana Eisenberg, and Hannah Wiseman for thoughtful suggestions and discussions, and to Brian Reaves and Matt Durose for their helpful corrections.

2. <i>Counsel for Indigent Defendants</i> .....	367
3. <i>Debtors' Prisons</i> .....	368
C. <i>Bulk Data and Reform</i> .....	370
III. EXISTING CRIMINAL JUSTICE DATA SOURCES .....	372
A. <i>Data Sources</i> .....	372
1. <i>State and Local Governments</i> .....	372
2. <i>The Federal Government</i> .....	379
3. <i>Commercial Services</i> .....	384
B. <i>Inadequate Incentives for Data Production</i> .....	386
1. <i>A Public Good Without a Market Solution</i> .....	387
2. <i>Local Opposition to Data Collection and Disclosure</i> .....	389
IV. CLOSING THE DATA GAP .....	390
A. <i>Improved Data Collection System</i> .....	391
B. <i>Federal Involvement</i> .....	393
V. OBJECTIONS .....	398
A. <i>Costs</i> .....	398
B. <i>Privacy Intrusions</i> .....	399
VI. CONCLUSION .....	399

## I. INTRODUCTION

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.”<sup>1</sup> If Google can digitize whole libraries of books<sup>2</sup> and firms can track consumers’ every purchase and preference,<sup>3</sup> the project of maintaining decent criminal justice statistics and posting them online seems elementary. Indeed, big data—the collection and analysis of large amounts of raw information to identify or understand phenomena in a variety of fields<sup>4</sup>—has

<sup>1</sup> VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA* 6 (2013).

<sup>2</sup> See *id.* at 83–84 (describing Google’s project, announced in 2004, to scan and digitize every book, and observing that by 2012, Google had electronically captured “more than 15 percent of the world’s written heritage”).

<sup>3</sup> Rory Van Loo, *Helping Buyers Beware: The Need for Supervision of Big Retail*, 163 U. PA. L. REV. 1311, 1331 (2015) (describing how firms “collect and mine data from billions of transactions” and observe consumer preferences by filming consumers’ movements and the direction of their eye movements as they shop within stores).

<sup>4</sup> The term “big data” describes a variety of types of data collection and analysis. See, e.g., MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 6 (noting that in the 2000s, “[t]he sciences like astronomy and genomics . . . coined the term,” observing that “[t]here is no rigorous definition of big data,” and defining the term as “things one can do at a large scale that cannot be done at a smaller one, to extract new insights or create new forms of value,

substantially changed how researchers, marketers, and policymakers approach problems. Thousands of terabytes<sup>5</sup> of electronic data have been scraped, coded, and run through countless statistical programs<sup>6</sup> to identify and analyze trends and suggest a targeted fix, such as improving the operation of the electrical grid, identifying dangerous drug interactions, and reducing traffic congestion.<sup>7</sup> More than a decade ago, Professor Daniel Esty noted how electronic data had already begun to transform environmental decision making, with electronic technologies making the collection, analysis, and dissemination of emissions and toxicology data less costly and far more effective.<sup>8</sup>

Twelve years and much technological progress later, the same cannot be said for crucial aspects of the criminal justice system. Collection and analysis of bulk data has been employed in parts of the system. Police now use aggregated law enforcement databases, extensive monitoring equipment, and

---

in ways that change markets, organizations, the relationship between citizens and governments, and more”); Michael Mattioli, *Disclosing Big Data*, 99 MINN. L. REV. 535, 539–42 (2014) (explaining that the term “refers to a new method of empirical inquiry,” including “practices that become more useful as electronic data recorded from devices and services grows,” and that big data allows the development of hypotheses after, rather than before, the collection of empirical evidence); Paul Ohm, Response, *The Underwhelming Benefits of Big Data*, 161 U. PA. L. REV. ONLINE 339, 339–40 (2013) (describing big data as “the trendy moniker for powerful new forms of data analytics,” and noting that the term has “become nearly synonymous with ‘data analysis’”); Neil M. Richards & Jonathan H. King, *Three Paradoxes of Big Data*, 66 STAN. L. REV. ONLINE 41, 42 (2013) (“Big data analytics depend on small data inputs, including information about people, places, and things collected by sensors, cell phones, click patterns, and the like. These small data inputs are aggregated to produce large datasets which analytic techniques mine for insight.”).

<sup>5</sup> See, e.g., MAYER-SCHÖNBERGER & CUKIER, *supra* note 1, at 9 (noting that in 2013 there were approximately 1,200 exabytes of stored information, only 2% of which was not digitized).

<sup>6</sup> See, e.g., *id.* at 2 (describing how Google explored 50 million common search terms used by Americans in order to identify areas of likely flu outbreak); *id.* at 4 (describing a big data project in which the researcher scraped 12,000 airline ticket prices from a 41-day period).

<sup>7</sup> *Id.* at 2 (describing a Google big data study and resulting model that appeared to be superior to the Center for Disease Control’s techniques in terms of identifying areas of flu outbreak); Mattioli, *supra* note 4, at 540–44 (describing a big data drug interaction study); Omer Tene & Jules Polonetsky, *Big Data for All: Privacy and User Control in the Age of Analytics*, 11 NW. J. TECH. & INTELL. PROP. 239, 245–51 (2013) (providing anecdotal accounts of these benefits and other applications). *But see* Ohm, *supra* note 4, at 340–43 (observing that it does not appear that Google’s flu outbreak tool has been applied in a beneficial manner—such as to prevent individuals from traveling to areas with full outbreaks—and that many other supposed big data successes lack proof of beneficial application).

<sup>8</sup> Daniel C. Esty, *Environmental Protection in the Information Age*, 79 N.Y.U. L. REV. 115, 156, 162 (2004).

complex models to project where and when crimes will occur.<sup>9</sup> But the benefits of data mining remain largely inaccessible to criminal justice reformers, policymakers, and academics—as well as judges and police that make the bulk of the decisions within the criminal justice system—because the data is not available to be mined. This is true even though the systems in place for arresting, detaining, trying, imprisoning, and releasing criminal defendants, in thousands of jurisdictions across the United States, are, in many ways, ideal targets for bulk data analytics. Millions of individuals are cycled through systems with common data points, such as the list of charges for each defendant; whether and when counsel were appointed, and why or why not; whether the individual was released on bail or held while awaiting trial; the length of pretrial imprisonment; the final disposition; and the terms of any sentence imposed. These types of data—which in many cases are already collected and recorded by some courts, but not electronically available in bulk format—could be anonymized and used to improve our understanding of the system as a whole.<sup>10</sup>

A recent in-depth analysis of Florida sentencing by the Sarasota Herald-Tribune showing that African-Americans receive significantly longer sentences than white defendants for similar crimes demonstrates both the potential of big data approaches and the limitations of existing datasets.<sup>11</sup> The paper’s investigative team reviewed “tens of millions of documents contained in two public record databases” in order to “compare sentences handed down by judges based on the points that defendants scored under Florida’s sentencing guidelines.”<sup>12</sup> As the paper explained, however, collecting the necessary data was an onerous task:

It took editors and reporters years to learn how to use the Offender Based Transaction System, and we only got to that point with help from the Tampa Bay Times.

---

<sup>9</sup> Michael Endler, *NYPD, Microsoft Push Big Data Policing into Spotlight*, DARK READING (Aug. 20, 2012), [http://www.darkreading.com/risk-management/nypd-microsoft-push-big-data-policing-into-spotlight/d/d-id/1105879?page\\_number=1](http://www.darkreading.com/risk-management/nypd-microsoft-push-big-data-policing-into-spotlight/d/d-id/1105879?page_number=1) [<https://perma.cc/8MGD-ZSYU>] (describing New York City’s Domain Awareness System); Mark van Rijmenam, *The Los Angeles Police Department Is Predicting and Fighting Crime with Big Data*, DATAFLOQ (Apr. 14, 2016), <https://datafloq.com/read/los-angeles-police-department-predicts-fights-crim/279> [<https://perma.cc/CB5S-ZHHW>]. For descriptions of similar efforts by police departments around the country, see Elizabeth E. Joh, *Policing by Numbers: Big Data and the Fourth Amendment*, 89 WASH. L. REV. 35, 42–48 (2014).

<sup>10</sup> Cf. Esty, *supra* note 8, at 132 (noting the value of bulk data in terms of identifying problems and that “[i]n the environmental domain there is often a great deal of uncertainty about whether a problem even exists”).

<sup>11</sup> Josh Salman et al., *Florida’s Broken Sentencing System*, SARASOTA HERALD-TRIB. (Dec. 11, 2016), <http://projects.heraldtribune.com/bias/sentencing/> [<https://perma.cc/AW4C-P66G>].

<sup>12</sup> Michael Braga, *How We Did It*, SARASOTA HERALD-TRIB., <http://projects.heraldtribune.com/bias/how-we-did-it/> [<https://perma.cc/CG8U-FTJS>].

Compiled by Florida’s court clerks, OBTS is a giant and unwieldy repository that tracks every criminal case from arrest through appeal.

.....

OBTS contains multiple entries for each defendant and crime, and all the duplicate entries had to be removed before meaningful analysis was possible.

After scrubbing the database, we were ready to start reporting in April. But it wasn’t long before we ran into another complication.

.....

Defendants . . . might receive longer sentences because of more prior convictions or they carried guns during robberies.

OBTS allows the user to correct for that weakness. It contains a field called “Sentence Special,” which identifies cases involving aggravating factors — such as extensive criminal histories and the use of a gun — that lead to longer sentences. We eliminated all such cases.

But court clerks do not enter the data in the “Sentence Special” field the same way in every county. So it is not always possible to get clean comparisons between two people charged with the same crime.

The Herald-Tribune overcame that limitation by using a second set of records compiled by the Florida Department of Corrections.<sup>13</sup>

The Herald-Tribune’s effort is impressive, but the problems with data access and availability its researchers had to overcome are both typical and daunting. Thus, scholars, advocacy groups, and government officials often rely on survey data from the Bureau of Justice Statistics.<sup>14</sup> These surveys, while extremely valuable, are limited in scope and frequency.<sup>15</sup> As a result, we do not know exactly how many inmates are detained at any given time, what they were charged with, the length of time between the charge and a verdict or plea, and numerous other data points that one might assume are readily available.<sup>16</sup> Improving the quality of—and access to—bulk data could thus open new avenues of research like that conducted by the Herald-Tribune, but bulk data could also be used to identify likely systemic constitutional violations at the

---

<sup>13</sup> *Id.*

<sup>14</sup> *See, e.g., infra* notes 50–51 and accompanying text.

<sup>15</sup> As to frequency, *see, for example, Data Collection: State Court Processing Statistics (SCPS)*, BUREAU JUST. STAT., [https://www.bjs.gov/index.cfm?ty=dcdetail&iid=282#Publications\\_and\\_products](https://www.bjs.gov/index.cfm?ty=dcdetail&iid=282#Publications_and_products) [<https://perma.cc/RTY3-X9GM>], noting that data was collected biennially between 1988 and 2006 and then again in 2009, which is the last survey for which data is available. As to scope, *see infra* note 16.

<sup>16</sup> As discussed *infra* Part II, the Bureau of Justice Statistics (BJS) collects extensive data on the number of individuals held in jail or supervised outside of jail, inmate mortality, and other data, although it often provides this data as a national compilation. *See, e.g.,* TODD D. MINTON & ZHEN ZENG, U.S. DEP’T OF JUSTICE, NCJ 248629, JAIL INMATES AT MIDYEAR 2014, at 1–9 (June 2015), <http://www.bjs.gov/content/pub/pdf/jim14.pdf> [<https://perma.cc/2L8D-2NTX>]. BJS also sometimes provides data on a state-specific but not county-specific level. *See infra* notes 49, 91, 94, and 166–67 and accompanying text (discussing other Bureau of Justice Statistics surveys and survey-based reports).

local level.<sup>17</sup> A jurisdiction, for example, with a high poverty rate but a low rate of appointed counsel may be violating the Sixth Amendment, while a jurisdiction that indiscriminately imposes burdensome conditions of pretrial release may run afoul of the Eighth and Fourteenth Amendments.<sup>18</sup>

Recent news reports,<sup>19</sup> lawsuits,<sup>20</sup> and scholarship<sup>21</sup> call attention to seemingly widespread violations of well-established constitutional rights

---

<sup>17</sup> Cf. Esty, *supra* note 8, at 133–34 (noting that, in the environmental context, the ability of bulk electronic data and analysis to identify the cause of environmental problems and the impact of these problems on particular populations or environments).

<sup>18</sup> See, e.g., *Walker v. City of Calhoun*, No. 4:15-CV-0170-HLM, 2016 WL 361612, at \*10 (N.D. Ga. Jan. 28, 2016) (“Any bail or bond scheme that mandates payment of pre-fixed amounts for different offenses to obtain pretrial release, without any consideration of indigence or other factors, violates the Equal Protection Clause.”), *vacated by*, No. 16-10521, 2017 WL 929750 (11th Cir. Mar. 9, 2017); *Thompson v. Moss Point*, No. 1:15cv182LG-RHW, 2015 WL 10322003, at \*1 (S.D. Miss., Nov. 6, 2015) (order granting declaratory judgment) (“No person may, consistent with the Equal Protection Clause of the Fourteenth Amendment . . . , be held in custody after an arrest because the person is too poor to post a monetary bond.”); *Jones v. City of Clanton*, No. 2:15cv34-MHT, 2015 WL 5387219, at \*1 (M.D. Ala. Sept. 14, 2015) (finding that jailing someone “because she was too poor to pay a small amount of bail money” violated the First, Eighth, and Fourteenth Amendments); *Pierce v. City of Velda City*, No. 4:15-cv-570-HEA, 2015 WL 10013006, \*2 (E.D. Mo. June 3, 2015) (describing a settlement, in which a city promised that “no person will be denied pretrial release because of their individual inability to make a monetary payment,” which arose as a result of an Equal Protection lawsuit).

<sup>19</sup> See, e.g., Jim Dwyer, *A Life That Frayed as Bail Reform Withered*, N.Y. TIMES (June 9, 2015), <http://www.nytimes.com/2015/06/10/nyregion/after-a-shocking-death-a-renewed-plea-for-bail-reform-in-new-york-state.html> [<https://perma.cc/LPP9-2GER>] (describing the death of Kalief Browder, who was jailed pretrial, for a long period of time, for a nonviolent crime that he repeatedly stated he did not commit); Sarah Stillman, *Get Out of Jail, Inc.*, NEW YORKER (June 23, 2014), <http://www.newyorker.com/magazine/2014/06/23/get-out-of-jail-inc> [<https://perma.cc/2HNT-GLNJ>] (describing the problem of debtors’ prison).

<sup>20</sup> *Walker*, 2016 WL 361612, at \*10; *Jones*, 2015 WL 5387219, at \*1 (showing class action and individual lawsuits addressing pretrial detention of defendants who could not afford money bail); Verified Complaint at 6–7, *Miller v. Deal*, No. 2011CV198121 (Ga. Super. Ct. Mar. 22, 2011).

<sup>21</sup> See, e.g., Lauren Sudeall Lucas, *Reclaiming Equality to Reframe Indigent Defense Reform*, 97 MINN. L. REV. 1197, 1198 (2013) (documenting the continuing, widespread failure to provide counsel to indigent defendants and relying primarily on studies that conducted surveys and collected data to provide average estimates of expenditures on indigent defense but did not document the number of defendants who qualify for but do not receive indigent defense); Samuel R. Wiseman, *Pretrial Detention and the Right to Be Monitored*, 123 YALE L.J. 1344, 1352–54 (2014) (describing the frequent jailing of nondangerous defendants pretrial using the available statistics, which are primarily produced from Bureau of Justice Statistics surveys rather than bulk, raw data and systemic analysis of those data); Ann K. Wagner, Comment, *The Conflict over Bearden v. Georgia in State Courts: Plea Bargained Probation Terms and the Specter of Debtors’ Prison*, 2010 U. CHI. LEGAL F. 383, 384–85 (describing the problem of debtors’ prison). For the studies, see, for example, NAT’L RIGHT TO COUNSEL COMM., THE CONSTITUTION PROJECT, JUSTICE DENIED: AMERICA’S CONTINUING NEGLECT OF OUR CONSTITUTIONAL RIGHT TO COUNSEL

within the system—both before, during, and after trial—and all have had to rely primarily on anecdotal evidence. “Debtors’ prisons,” in which individuals are imprisoned for a failure to pay court costs associated with probation, child support, or other required fees, appear to be a common phenomenon,<sup>22</sup> but identifying the most problematic individual jurisdictions remains a largely local endeavor.<sup>23</sup> Further, it appears that indigent individuals in some jurisdictions are routinely denied counsel at critical stages;<sup>24</sup> this is a problem that could be identified by sifting through docket-level data, but because much of the data is not available electronically, it must be discovered through haphazard (and costly) local investigation.<sup>25</sup> In the bail context, the death of Kalief Browder, while awaiting trial for a crime that he did not commit, called attention to the months—even years—individuals who cannot afford bail sit in jail awaiting trial.<sup>26</sup> Yet, we do not know which jurisdictions hold individuals for the longest periods of time, or which jurisdictions tend to imprison the most defendants simply because these defendants, many of whom are not dangerous, cannot afford the bail amount imposed.<sup>27</sup> As a result, in part, of

---

85–89 (Apr. 2009), <http://www.constitutionproject.org/wp-content/uploads/2012/10/139.pdf> [<https://perma.cc/M945-2GKC>], which provides examples of indigent defendants for whom a court failed to appoint counsel or appointed counsel too late but does not indicate the percentage or number of defendants nationwide for whom counsel was denied or provided too late.

<sup>22</sup> See, e.g., Stillman, *supra* note 19 (providing anecdotal accounts of the problem).

<sup>23</sup> See, e.g., Verified Complaint, *supra* note 20, at 5 (regarding individuals who were jailed for a failure to pay child support in a class action, relying largely on news articles to document these cases of debtors’ prison); HUMAN RIGHTS WATCH, PROFITING FROM PROBATION: AMERICA’S “OFFENDER-FUNDED” PROBATION INDUSTRY 10–11 (2014), [https://www.hrw.org/sites/default/files/reports/us0214\\_ForUpload\\_0.pdf](https://www.hrw.org/sites/default/files/reports/us0214_ForUpload_0.pdf) [<https://perma.cc/67AB-KFZW>] (relying primarily on seventy-five interviews, visits to county and municipal courts and open records requests sent to these courts, media and literature reviews, and conversations with private probation firms to document the problem of debtors’ prison).

<sup>24</sup> See, e.g., Mary Sue Backus & Paul Marcus, *The Right to Counsel in Criminal Cases, a National Crisis*, 57 HASTINGS L.J. 1031, 1031–36 (2006) (collecting anecdotes of counsel appointed too late in cases around the country and describing numerous reports’ conclusions that indigent defendants are not consistently receiving adequate defense).

<sup>25</sup> *Rothgery v. Gillespie County*, 554 U.S. 191, 195–96, 213 (2008) (describing as “routine” a hearing required by Texas law in which the defendant is “formally apprised” of the charges against him, bail is set, and the defendant is jailed if he cannot afford bail, and how a county did not provide counsel to indigent defendants at this stage—a practice that the Supreme Court determined was unconstitutional); *Heckman v. Williamson County*, 369 S.W.3d 137, 156–57 (Tex. 2012) (describing a routine failure to appoint counsel for indigent defendants).

<sup>26</sup> See Dwyer, *supra* note 19.

<sup>27</sup> The available statistics in this area tend to be national statistics based on Bureau of Justice Statistics surveys and estimates by other national groups. See, e.g., THOMAS H. COHEN & BRIAN A. REAVES, U.S. DEP’T OF JUSTICE, NCJ 214994, PRETRIAL RELEASE OF FELONY DEFENDANTS IN STATE COURTS I (Nov. 2007), <http://bjs.ojp.usdoj.gov/content/pu>

these types of data limitations, the Justice Department's Civil Rights Division had to conduct a twenty-month investigation to identify apparent racial bias in Ferguson, Missouri's juvenile justice system.<sup>28</sup> Knowledge of systemic problems in similar systems around the country remains underdeveloped.<sup>29</sup>

The potential applications for improved criminal justice data are numerous, and there is increasing recognition of the importance of data and statistics in criminal justice reform efforts. The President's 2015 budget included improved collection of information on "public defender agencies, programs and operations," recidivism rates, inmates in prisons and jails, and state criminal data, among other types of information.<sup>30</sup> The President's Task Force on 21st Century Policing has launched a "Police Data Initiative," in which twenty-one cities and counties are attempting to improve and unify the data collected and recorded through many different computer systems in each jurisdiction to support better analysis of problems and develop policing "best practices."<sup>31</sup> Further, the Federal Bureau of Justice Statistics collects data on defendant deaths in custody, including deaths during arrests, although it is only voluntarily provided by law enforcement agencies and thus only available for limited jurisdictions.<sup>32</sup> The goal is to "gain a national picture of police use of force" and to "incentivize the systematic and transparent collection and analysis of use of force incident data at the local level."<sup>33</sup>

Nonetheless, the development of usable data, particularly post-arrest data, has been slow. The failure to make available these basic, widely collected data

---

b/pdf/prfdsc.pdf [https://perma.cc/75ZX-PNCK] (noting that five out of six defendants detained before trial could not afford the bail set in their case, based on a survey of state courts and pretrial services agencies in forty of the seventy-five largest counties); CRIMINAL JUSTICE SECTION, AM. BAR ASS'N, STATE POLICY IMPLEMENTATION PROJECT 2, [http://www.americanbar.org/content/dam/aba/administrative/criminal\\_justice/spip\\_handout\\_s.authcheckdam.pdf](http://www.americanbar.org/content/dam/aba/administrative/criminal_justice/spip_handout_s.authcheckdam.pdf) [https://perma.cc/QWE8-AJZW] ("Two thirds of the 500,000 individuals incarcerated in jail and awaiting trial are low bail risk, meaning they have been deemed by a magistrate to pose no significant risk to themselves or the community, as well as representing a low risk of flight.").

<sup>28</sup>CIVIL RIGHTS DIV., U.S. DEP'T OF JUSTICE, INVESTIGATION OF THE ST. LOUIS COUNTY FAMILY COURT ST. LOUIS, MISSOURI (July 2015), <http://www.justice.gov/opa/file/641971/download> [https://perma.cc/U92K-55J7].

<sup>29</sup>OFFICE OF MGMT. & BUDGET, EXEC. OFFICE OF THE PRESIDENT OF THE U.S., FISCAL YEAR 2015: ANALYTICAL PERSPECTIVES, BUDGET OF THE U.S. GOVERNMENT 294 (2015), <https://www.gpo.gov/fdsys/pkg/BUDGET-2015-PER/pdf/BUDGET-2015-PER.pdf> [https://perma.cc/SS2F-GRKP].

<sup>30</sup>*Id.*

<sup>31</sup>Megan Smith & Roy L. Austin, Jr., *Launching the Police Data Initiative*, WHITE HOUSE (May 18, 2015), <https://www.whitehouse.gov/blog/2015/05/18/launching-police-data-initiative> [https://perma.cc/YKL8-HWFW].

<sup>32</sup>PRESIDENT'S TASK FORCE ON 21ST CENTURY POLICING, FINAL REPORT 21–22 (May 2015), <http://www.policemag.com/resources/documents/21stcenturytaskforce-finalreport.pdf> [https://perma.cc/QCL7-78FV].

<sup>33</sup>*Id.* at 22.



sources arises, in large part, from a lack of adequate incentives.<sup>34</sup> Unlike the law firms that pay Westlaw and LexisNexis large sums to access those databases, and that collect docket-level data on their own,<sup>35</sup> the entities that would use these statistics tend to be underfunded nonprofit groups,<sup>36</sup> indigent criminal defendants,<sup>37</sup> and academics.<sup>38</sup> Individual jurisdictions, moreover, may be hesitant to expose the workings of their court systems in order to avoid unwanted attention.<sup>39</sup> Criminal justice data, then, is an underproduced public good,<sup>40</sup> and further<sup>41</sup> government intervention is needed. The federal government, or individual states operating under federal guidance for the uniform production and reporting of information, must step in to incentivize or mandate the production and online posting of basic criminal justice data.

This Article identifies the problem of the black box of criminal justice data,<sup>42</sup> explores the reasons behind it, and proposes solutions. Part I explores

<sup>34</sup> See *infra* notes 35–41 and accompanying text.

<sup>35</sup> See *infra* Part III.A.

<sup>36</sup> See, e.g., AM. CIVIL LIBERTIES UNION, IN FOR A PENNY: THE RISE OF AMERICA'S NEW DEBTORS' PRISONS 6–8 (Oct. 2010), [https://www.aclu.org/sites/default/files/field\\_document/InForAPenny\\_web.pdf](https://www.aclu.org/sites/default/files/field_document/InForAPenny_web.pdf) [<https://perma.cc/4RC7-Z9HB>] (describing problems in a variety of states based on surveys and interviews); *Ending the American Money Bail System*, EQUAL JUST. UNDER L., <http://equaljusticeunderlaw.org/wp/current-cases/ending-the-american-money-bail-system/> [<https://perma.cc/59KM-33DY>] (describing class action lawsuits filed by the group in U.S. district courts in an attempt to end money bail on a case-by-case basis).

<sup>37</sup> See, e.g., *Rothgery v. Gillespie County*, 554 U.S. 191, 212–13 (2008) (reviewing an indigent defendant's challenge to the denial of counsel by a county court at a critical stage); Lucas, *supra* note 21, at 1199–200 (noting that to address the indigent defense crisis, “legal advocates have generally based litigation reform efforts on the Sixth Amendment” using post-conviction ineffective assistance of counsel claims or “civil class-action lawsuits”).

<sup>38</sup> See, e.g., Wiseman, *supra* note 21, 1351–63 (documenting the problematic effects of money bail and pretrial detention relying on nationally available summary data and limited examples from states and counties); Wagner, *supra* note 21, at 391–94 (documenting the problem of debtors' prisons largely through anecdotal evidence); see also Laura I. Appleman, *Justice in the Shadowlands: Pretrial Detention, Punishment, & the Sixth Amendment*, 69 WASH. & LEE L. REV. 1297, 1302 (2012) (same); Shima Baradaran, *The State of Pretrial Detention* (same), in *THE STATE OF CRIMINAL JUSTICE 2011*, at 187, 190 (Am. Bar Ass'n ed., 2011).

<sup>39</sup> See *infra* notes 244–49 and accompanying text.

<sup>40</sup> See *infra* notes 236–39 and accompanying text.

<sup>41</sup> As introduced above and discussed further below, the Bureau of Justice Statistics already collects large amounts of data in survey form from the seventy-five largest counties and the federal courts, as well as from public defender agencies, but this data is not provided on a jurisdiction-by-jurisdiction basis and lacks the types of data points needed to identify what this Article identifies as some of the most pressing criminal justice issues. See *supra* notes 32–33 and accompanying text; *infra* notes 104–06, 173–99 and accompanying text.

<sup>42</sup> This is the first Article to identify the criminal justice data “black box.” Other scholars have used this term to describe certain aspects of the system. See Pamela Metzger & Andrew Guthrie Ferguson, *Defending Data*, 88 S. CAL. L. REV. 1057, 1059 (2015) (“[P]ublic defense remains a largely data-less enterprise, a black box of discretionary

the areas in which bulk criminal justice data would be most valuable and the groups that could benefit from its availability, describing how academics, policymakers, nonprofit groups, judges, and police could identify systemic violations of well-established constitutional rights in the criminal justice system and focus improvement efforts on jurisdictions with the highest violation of rights. Part II then documents the current gaps in available information and explores the factors that likely lead to the underproduction of data.

After identifying the problem and its likely causes, Part III of this Article calls for an improved data collection system in which county and state dockets contain basic, uniform information points, such as demographic data including the age, race, and gender of the defendant; the type of charge; date of the charge; whether or not counsel was appointed, date of appointment, and reason for the appointment or failure to appoint; date of the bail hearing; pretrial release conditions imposed if the defendant was released; days spent in pretrial detention if the defendant was jailed; whether or not the defendant appeared for trial or committed crimes pretrial; hearings scheduled and held; and the type of adjudication (plea, trial, or diversion, for example). Part III also explores models from which this data collection system could build, analyzing the benefits and drawbacks of existing state and federal efforts to unify criminal justice data collection.

Finally, in Part IV, this Article identifies and responds to likely objections, including the cost of collecting and posting online criminal justice data for millions of defendants and concerns about defendants' privacy. The fact that many dockets already contain much of the information that this Article calls for, and that many are electronically available in courthouses, yet are simply not posted online, suggests that the improved system would not be prohibitively costly.<sup>43</sup> Further, much of the data currently provided for a fee from the service Public Access to Court Electronic Records (PACER)<sup>44</sup> does not appear to require the funds generated by PACER. Courts often use PACER revenues to fund courtroom technologies and other activities unrelated to the dissemination of information about county and state cases.<sup>45</sup> Further, while privacy concerns are legitimate and must be carefully addressed, data entry systems could be designed to exclude or redact defendant names, addresses, and other personal identifying characteristics,<sup>46</sup> and the computer systems

---

decisions disconnected from any systemic analysis about the relationship between defender practices and case outcomes.”)

<sup>43</sup> See PACER, <https://www.pacer.gov/> [<https://perma.cc/ZZF8-ZRAD>].

<sup>44</sup> *Id.*

<sup>45</sup> Stephen Schultze, *What Does It Cost to Provide Electronic Public Access to Court Records?*, MANAGING MIRACLES (May 29, 2010), <http://managingmiracles.blogspot.com/2010/05/what-is-electronic-public-access-to.html> [<https://perma.cc/GRQ8-69YZ>]; *infra* note 242 and accompanying text.

<sup>46</sup> See, e.g., *Idaho Supreme Court Data Repository*, IDAHO ST. JUDICIARY, <https://www.idcourts.us/repository/start.do> [<https://perma.cc/5Y5U-NDES>] (describing

could automatically screen for and delete, or otherwise black out, these types of data as a backstop measure.

## II. THE USES OF BULK CRIMINAL JUSTICE DATA

The empirical analysis of bulk criminal justice data has produced promising results in the limited contexts where it can currently be employed,<sup>47</sup> but broader applications are hampered by a lack of access to the necessary information. The American criminal justice system involves thousands of local, state, and federal officials who arrest, detain, or release on bail, imprison, sentence, and establish parole conditions for millions of defendants. This multiplicity of systems makes collecting and analyzing uniform data across jurisdictions difficult—but, due to the dramatic, technology-driven fall in the costs of storing and disseminating information, no longer an insuperable challenge. Making existing information from local, state, and federal courts' criminal dockets and case summaries electronically available and standardizing data points within dockets to ensure uniformity could trigger meaningful change in the criminal justice system at both the national and local levels. This Part discusses the substantive areas of the system where improved data and analysis could have significant impacts and the ways in which the data could be used by different groups to study national issues and identify problems at the state and local levels.

### A. Data and Systemic Problems

Currently, criminal justice researchers have to either rely on limited survey data or engage in the costly, cumbersome task of collecting and coding more complete data sets from a patchwork of federal, state, and local systems.<sup>48</sup> In the absence of available bulk data, researchers frequently rely on sample data from the Bureau of Justice Statistics (BJS). Professors Shima Baradaran and Frank McIntyre, for example, studied 116,000 observations of felony defendants from BJS State Court Processing Statistics in forty<sup>49</sup> out of the seventy-five largest U.S. counties over a sixteen-year period.<sup>50</sup> They included numerous data points for each defendant, such as “initial crime committed, any subsequent bail crime, and the prior record of the defendant, including any

---

Idaho courts' omissions of personal identifying data in their electronic system that collects criminal defendant arrest, bail, and case disposition information).

<sup>47</sup> See *infra* notes 49–50, 59–61, 67–68 and accompanying text.

<sup>48</sup> See *supra* notes 9–29 and accompanying text.

<sup>49</sup> *State Court Processing Statistics, 1990-2009: Felony Defendants in Large Urban Counties (ICPSR 2038)*, ICPSR, <http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/2038> [<https://perma.cc/4X7C-AC72>].

<sup>50</sup> Shima Baradaran & Frank L. McIntyre, *Predicting Violence*, 90 TEX. L. REV. 497, 524–25 (2012).

failures to appear in court,”<sup>51</sup> demographic characteristics, and pretrial crimes committed. They then used this and other data to construct a model for better predicting the dangerousness of defendants based on certain defendant data, such as the prior record.<sup>52</sup> Professors Baradaran and McIntyre concluded that, if judges used their model for predicting dangerousness and thus for determining who should be detained pretrial, 25% more defendants could be released than are currently released, and pretrial crime rates would decline.<sup>53</sup>

But survey data is limited both in content and scope. The data from the forty counties used in the Baradaran-McIntyre study includes only a limited time window within each survey year,<sup>54</sup> and it does not include information such as each “defendant’s community ties, employment status, income . . . and mental health status” due to “cost and data accessibility.”<sup>55</sup> Further, the system “cannot distinguish defendants released under conditions that involve intensive pretrial monitoring from defendants released under less stringent pretrial conditions.”<sup>56</sup> And raw BJS data is not readily available online. Researchers are required to obtain approval from their supervising institution and work with the BJS to obtain it.<sup>57</sup> Researchers who need data not collected by the BJS must either give up or attempt the time-consuming task of assembling their own datasets.<sup>58</sup>

In an effort somewhat similar to McIntyre’s and Baradaran’s—but which involved somewhat more detailed digging into primary data—the Laura and John Arnold Foundation hired leading criminal justice researchers to identify “1.5 million cases drawn from more than 300 U.S. jurisdictions,” a dataset that was then narrowed to 746,525 defendants that had been released pretrial.<sup>59</sup> These researchers then identified combined information about these

---

<sup>51</sup> *Id.* at 525 (“Every two years, the ten largest U.S. counties are automatically surveyed, as are thirty other counties drawn from the next sixty-five largest counties.”).

<sup>52</sup> *Id.* at 528.

<sup>53</sup> *Id.* at 558.

<sup>54</sup> BUREAU OF JUSTICE STATISTICS, U.S. DEP’T OF JUSTICE, DATA ADVISORY: STATE COURT PROCESSING STATISTICS DATA LIMITATIONS (Mar. 2010), [http://www.bjs.gov/content/pub/pdf/scpsdl\\_da.pdf](http://www.bjs.gov/content/pub/pdf/scpsdl_da.pdf) [<https://perma.cc/6Z5P-U6NV>].

<sup>55</sup> *Id.*

<sup>56</sup> *Id.*

<sup>57</sup> See *State Court Processing Statistics, 1990-2009: Felony Defendants in Large Urban Counties (ICPSR 2038)*, *supra* note 49 (showing “[n]o downloadable data files available” for datasets and requiring an online application for data access). The BJS has good reasons for not posting all of the raw data, including concerns about defendant confidentiality and concerns that data users could draw the wrong conclusions from data and publish misleading results. Indeed, the BJS is careful to emphasize the lessons that cannot be produced from its data, noting that the data cannot show “the efficacy of one form of pretrial release over another” or the “effectiveness of a particular program in preventing pretrial misconduct.” BUREAU OF JUSTICE STATISTICS, *supra* note 54.

<sup>58</sup> See, e.g., *infra* notes 59–80 and accompanying text.

<sup>59</sup> LJAF, DEVELOPING A NATIONAL MODEL FOR PRETRIAL RISK ASSESSMENT 3 (Nov. 2013), [http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary\\_PSA-Court\\_4\\_1.pdf](http://www.arnoldfoundation.org/wp-content/uploads/2014/02/LJAF-research-summary_PSA-Court_4_1.pdf) [<https://perma.cc/R7VZ-37LS>].

defendants, such as prior criminal records, with information on whether the defendants committed a crime while released before trial or failed to appear before trial, studying “hundreds of risk factors” that might best predict flight and pretrial crime.<sup>60</sup> They ultimately identified nine risk factors that had the strongest predictive force,<sup>61</sup> and the foundation used this information to create a nationally applicable model that all jurisdictions could use to best predict risk and to avoid detaining defendants with low risk factors.<sup>62</sup> All counties in Kentucky began using the foundation’s model in 2013, and in the first six months of use, the state increased its release of defendants from 68% of defendants released to 70%, and pretrial crime declined by 15%.<sup>63</sup> The foundation also believes that the Kentucky results so far show that the tool is “racially neutral and gender neutral” in that “[i]t accurately classifies defendants’ risk levels regardless of their race or gender.”<sup>64</sup> Since Kentucky’s adoption of the tool, more than nineteen other cities, counties, and states have adopted the tool, including some of the most populous cities.<sup>65</sup> This laudable effort demonstrates the potential for data analysis to improve criminal justice, but it also shows how daunting the task of collecting criminal justice data can be. This effort would not have been possible without the generous funding of the foundation, which was formed by a billionaire couple.<sup>66</sup>

In the habeas context, Professor Nancy King and two experts with the National Center for State Courts conducted an extensive empirical analysis of habeas litigation in U.S. district courts.<sup>67</sup> The goal of the authors was to identify success rates in habeas corpus cases filed by state prisoners under the Antiterrorism and Effective Death Penalty Act of 1996, with a purpose of informing policymakers, litigants, courts, and scholars.<sup>68</sup> State prisoners file more than 18,000 habeas cases annually, and approximately 6,000 of these

---

<sup>60</sup> *Id.*

<sup>61</sup> *Id.* at 3–4.

<sup>62</sup> *Id.* at 4.

<sup>63</sup> *More than 20 Cities and States Adopt Risk Assessment Tool to Help Judges Decide Which Defendants to Detain Prior to Trial*, LJAF (June 26, 2015), <http://www.arnoldfoundation.org/more-than-20-cities-and-states-adopt-risk-assessment-tool-to-help-judges-decide-which-defendants-to-detain-prior-to-trial/> [<https://perma.cc/9Q9W-MMCW>].

<sup>64</sup> LJAF, RESULTS FROM THE FIRST SIX MONTHS OF THE PUBLIC SAFETY ASSESSMENT – COURT IN KENTUCKY 4 (July 2014) (emphasis omitted), <http://www.arnoldfoundation.org/wp-content/uploads/2014/02/PSA-Court-Kentucky-6-Month-Report.pdf> [<https://perma.cc/YB7Y-6RWK>].

<sup>65</sup> *More than 20 Cities and States Adopt Risk Assessment Tool to Help Judges Decide Which Defendants to Detain Prior to Trial*, *supra* note 63.

<sup>66</sup> Caroline Preston, *A Thirtysomething Billionaire Couple Take on Tough Issues via Giving*, CHRON. PHILANTHROPY (Oct. 16, 2011), <https://philanthropy.com/article/A-Thirtysomething-Billionaire/157613> [<https://perma.cc/LKB7-GGKS>].

<sup>67</sup> NANCY J. KING ET AL., FINAL TECHNICAL REPORT: HABEAS LITIGATION IN U.S. DISTRICT COURTS 7 (Aug. 2007), <https://www.ncjrs.gov/pdffiles1/nij/grants/219559.pdf> [<https://perma.cc/N2FL-43UR>].

<sup>68</sup> *Id.*

cases are heard by federal courts of appeals each year.<sup>69</sup> Yet, as King, Cheesman, and Ostrom note, there was little empirical information about these cases prior to their resource-intensive study, including information about the delays between a state court's habeas denial and a federal habeas petition; the types of habeas challenges raised, such as actual innocence, and whether prisoners tend to challenge their sentences, convictions, or both; the percentages of habeas cases that fail due to procedural problems such as a failure to exhaust claims in state court; the time that it takes for courts to process habeas cases; and case outcomes.<sup>70</sup> King, Cheesman, and Ostrom observe that datasets "maintained by the Administrative Office of the United States Courts" contain individual case statistics on "filing and termination dates, disposition information, whether the petitioner had counsel, and if the petitioner received in forma pauperis status," but "information about which party prevailed is missing for most habeas cases."<sup>71</sup>

To obtain more thorough data, the researchers used the federal PACER<sup>72</sup> database of case information, and because PACER information for cases from 2002 and earlier was inadequate, they limited their sample to cases filed in 2003 and 2004 for noncapital cases, selecting a random set of 7.5% of the cases to study.<sup>73</sup> After removing mislabeled and duplicate cases from the set, they studied a total of 2,384 noncapital cases from a diverse geographic area.<sup>74</sup> The researchers needed pre-2002 data for capital cases because these cases travel through the courts for so many years, and for these samples they obtained hard copy data, collecting "original case documents at courthouses and federal archives"; they also used some online research for the capital dataset.<sup>75</sup> To pay for the use of PACER and the many researchers who helped gather and analyze the information, King and the other researchers obtained a National Institute of Justice grant.<sup>76</sup>

Perhaps most telling is Professor Brandon Garrett's experience researching the recent, national decline in death sentences for a forthcoming book.<sup>77</sup> Despite the enormous attention paid to the death penalty, there was no existing dataset containing jurisdiction-specific information on death sentences from the last twenty years.<sup>78</sup> Assembling the data required him to consult

---

<sup>69</sup> *Id.* at 9–10.

<sup>70</sup> *Id.* at 11–12.

<sup>71</sup> *Id.* at 13.

<sup>72</sup> PACER, *supra* note 43.

<sup>73</sup> KING ET AL., *supra* note 67, at 15.

<sup>74</sup> *Id.*

<sup>75</sup> *Id.*

<sup>76</sup> *Id.* at 2.

<sup>77</sup> BRANDON L. GARRETT, *END OF ITS ROPE: HOW THE DEMISE OF THE DEATH PENALTY CAN REVIVE CRIMINAL JUSTICE* (forthcoming 2017) (manuscript at 4) (on file with author) (cited with author's permission).

<sup>78</sup> *Id.*

“prison websites, death penalty appeals, and . . . death penalty lawyers around the county.”<sup>79</sup>

These efforts to identify nationwide criminal justice trends show the currently herculean task of collecting the necessary data and how a funding source, such as a foundation or government grant, is typically needed to complete such a large task. The collection of criminal justice data remains a labor-intensive effort, in contrast with other contexts in which large amounts of raw data are readily found on the internet.<sup>80</sup>

### B. *Data and Local Problems*

Bulk data could also be enormously useful in remedying criminal justice problems at the local level. Much legal scholarship focuses on persuading courts to expand constitutional rights.<sup>81</sup> But in many areas of criminal justice, there is a widespread belief that existing rights are underenforced.<sup>82</sup> The nature and extent of these violations are, however, difficult to establish in light of the largely (and necessarily, in the absence of more complete data) available anecdotal evidence.<sup>83</sup> If more comprehensive data were available, it could be used to identify likely local targets for reform litigation and legislation.<sup>84</sup>

Three areas of the criminal justice system where existing anecdotal and survey-based data suggest that well-established rights are routinely violated in some parts of the country—or where great inequities exist even if not amounting to a constitutional violation—include bail,<sup>85</sup> counsel for indigent

---

<sup>79</sup> *Id.*

<sup>80</sup> See *supra* notes 1–9 and accompanying text.

<sup>81</sup> See, e.g., Shima Baradaran, *Restoring the Presumption of Innocence*, 72 OHIO ST. L.J. 723, 768–76 (2011) (arguing for more extensive pretrial justice rights); Justin F. Marceau, *Don’t Forget Due Process: The Path Not (Yet) Taken in § 2254 Habeas Corpus Adjudications*, 62 HASTINGS L.J. 1, 46–47 (2010) (arguing that due process entitles criminal defendants to de novo review in federal court when they lacked a full and fair adjudication of their federal constitutional claims in state court).

<sup>82</sup> See, e.g., Wiseman, *supra* note 21, at 1351–63 (exploring the problem of money bail and, although documenting potential constitutional arguments against its use, also focusing on the policy reasons for changing existing practices).

<sup>83</sup> See, e.g., *supra* notes 22–24 and accompanying text; *infra* notes 222–24 and accompanying text.

<sup>84</sup> The lack of information on the workings of state criminal justice systems and the need to uncover persistent problems has caused some scholars to suggest creative workarounds. Eve Brensike Primus, for example, has suggested that habeas cases should be limited to “structural reform litigation” to address widespread, systemic abuses of constitutional rights during criminal convictions. Eve Brensike Primus, *A Structural Vision of Habeas Corpus*, 98 CALIF. L. REV. 1, 44 (2010); see also Esty, *supra* note 8, at 137 (noting the value of improved information collection, analysis, and dissemination to informing policy debates in the environmental context).

<sup>85</sup> See *infra* Part II.B.1.

defendants,<sup>86</sup> and imprisonment for inability to pay a court-ordered fine or fee, commonly referred to as debtors' prison.<sup>87</sup>

### 1. *Bail and Pretrial Detention*

In the bail context, many judges routinely set bail based on schedules with monetary amounts that reflect the type of charge, not the defendant's ability to pay.<sup>88</sup> This manner of setting bail—which is used to ensure defendants' presence at trial and good behavior pending trial—allows wealthy defendants, even those charged with relatively high-level crimes and who are potentially incentivized to flee, to pay their way out of jail while awaiting trial, and it causes large percentages of low-risk, nondangerous, indigent defendants to be detained before trial.<sup>89</sup>

Numerous scholarly articles, publications, and studies by nonprofit groups, and a small but growing number of class action cases, have begun to document the apparently widespread problems in this area.<sup>90</sup> The American Bar Association estimates that “[t]wo thirds of the 500,000 individuals incarcerated in jail and awaiting trial are low bail risk, . . . pos[ing] no significant risk to themselves or the community, as well as representing a low risk of flight.”<sup>91</sup> The Justice Policy Institute observes that “[m]oney determines pretrial release for 7 out of 10 people accused of felonies” and that “25 percent more people could be safely released from U.S. jails while awaiting trial if the proper procedures are put in place.”<sup>92</sup> And a small scholarly literature provides examples from several jurisdictions of the hardships pretrial detention imposes on defendants and their families, such as the loss of jobs and housing.<sup>93</sup>

---

<sup>86</sup> See *infra* Part II.B.2.

<sup>87</sup> See *infra* Part II.B.3.

<sup>88</sup> PRETRIAL JUSTICE INST., PRETRIAL JUSTICE IN AMERICA: A SURVEY OF COUNTY PRETRIAL RELEASE POLICIES, PRACTICES AND OUTCOMES 7 (Sept. 2010), <http://www.pretrial.org/wpfb-file/pji-pretrial-justice-in-america-a-survey-of-county-pretrial-release-policies-practices-and-outcomes-2010-pdf> [https://perma.cc/C6VT-6D5R] (showing one survey that indicates 64% of counties use bail schedules).

<sup>89</sup> See *Attorney General Eric Holder Speaks at the National Symposium on Pretrial Justice*, U.S. DEP'T JUST. (June 1, 2011), <http://www.justice.gov/opa/speech/attorney-general-eric-holder-speaks-national-symposium-pretrial-justice> [https://perma.cc/WBN7-DX5D] (noting that many defendants jailed pretrial are “nonviolent, non-felony offenders, charged with crimes ranging from petty theft to public drug use”).

<sup>90</sup> See *infra* notes 91–93 and accompanying text.

<sup>91</sup> CRIMINAL JUSTICE SECTION, *supra* note 27, at 2.

<sup>92</sup> JUSTICE POLICY INST., BAIL FAIL: WHY THE U.S. SHOULD END THE PRACTICE OF USING MONEY FOR BAIL 2, 4 (Sept. 2012), <http://www.justicepolicy.org/uploads/justicepolicy/documents/bailfail.pdf> [https://perma.cc/33YL-9ZYC].

<sup>93</sup> See, e.g., Appleman, *supra* note 38, at 1319–21; Baradaran, *supra* note 38, at 334; Douglas L. Colbert, *Prosecution Without Representation*, 59 BUFF. L. REV. 333, 334 (2011); Wiseman, *supra* note 21, at 1353–54.



To address some of these problems, Equal Justice Under Law (EJUL), a nonprofit group engaged in criminal justice reform litigation and advocacy,<sup>94</sup> has begun to identify some of the problematic jurisdictions. It has filed several lawsuits in U.S. district courts arguing that jailing indigent defendants solely on the basis of their inability to pay a small monetary fine—in this case a fine set by a fixed bail schedule—violates equal protection and due process.<sup>95</sup> Notably, the U.S. Department of Justice (DOJ) submitted a “Statement of Interest” in one of these cases, arguing that setting bail without any consideration of a defendant’s ability to pay is unconstitutional.<sup>96</sup> EJUL has achieved remarkable results, securing numerous rulings against the use of money bail without an inquiry into ability to pay.<sup>97</sup>

This type of litigation will always require on the ground investigation into local practices, but analysis of bulk data could help identify the jurisdictions likely engaging in unconstitutional practices. Indeed, the limited data available shows that release rates vary widely among jurisdictions (as well as demographic classes),<sup>98</sup> and it is currently difficult, if not impossible, to know which county, state, or federal jails tend to most frequently host defendants who are detained simply because they could not pay; which judges in which

---

<sup>94</sup> *About Us*, EQUAL JUST. UNDER L., <http://equaljusticeunderlaw.org/wp/about-us/> [<https://perma.cc/BBT6-GDQT>].

<sup>95</sup> *See, e.g.*, Class Action Complaint at 11, *Snow v. Lambert*, No.15-cv-00567 (M.D. La. Aug. 25, 2015) (“The Fourteenth Amendment’s due process and equal protection clauses have long prohibited keeping a person in jail because of the person’s inability to make a monetary payment. Defendants violate Plaintiff’s rights by placing and keeping her in jail when she cannot afford to pay the amount of money set by the generic secured bail ‘schedule’ used by Defendants.”).

<sup>96</sup> Statement of Interest of the United States, *Varden v. City of Clanton*, No. 2:15-cv-00034-MHT-WC (M.D. Ala. Feb. 13, 2015).

<sup>97</sup> *Ending the American Money Bail System*, *supra* note 36.

<sup>98</sup> *See, e.g.*, THOMAS H. COHEN, U.S. DEP’T OF JUSTICE, NCJ 239243, PRETRIAL RELEASE AND MISCONDUCT IN FEDERAL DISTRICT COURTS, 2008–2010, at 3 (Nov. 2012), <http://www.bjs.gov/content/pub/pdf/prmfdc0810.pdf> [<https://perma.cc/9X4D-85DA>] (noting an overall 36% pretrial release rate in federal district courts but not identifying different release rates for different courts); *id.* at 10 (noting large differences in pretrial release rates “between black and white defendants,” particularly for drug offenses, with 60% of white defendants released and 36% of black defendants released on average across the federal district courts); *Frequently Asked Questions About Pretrial Release Decision Making*, AM. BAR ASS’N., [http://www.americanbar.org/content/dam/aba/publications/criminaljustice/FAQ\\_Pretrial\\_Justice.authcheckdam.pdf](http://www.americanbar.org/content/dam/aba/publications/criminaljustice/FAQ_Pretrial_Justice.authcheckdam.pdf) [<https://perma.cc/4ULM-Q7VB>] (describing successful pretrial release programs, and noting 80% release rates in the District of Columbia and 74% release rates in Kentucky before Kentucky began using the Laura and John Arnold Foundation risk assessment, which further increased release); *cf.* OFFICE OF PROGRAM POLICY ANALYSIS & GOV’T ACCOUNTABILITY, No. 08-75, PRETRIAL RELEASE PROGRAMS VARY ACROSS THE STATE; NEW REPORTING REQUIREMENTS POSE CHALLENGES 9 (Dec. 2008), <http://www.oppaga.state.fl.us/reports/pdf/0875rpt.pdf> [<https://perma.cc/2668-J7UZ>] (showing the number of participants in pretrial release programs in Florida counties as ranging from 113 in Citrus County to 11,101 in Dade County).

jurisdictions tend to rely most heavily on bail schedules rather than on defendants' ability to pay; how often and in what manner released defendants violate the terms of their release; and how long each defendant is detained while awaiting trial. With more data available (and analyzed), reform groups and government agencies working in this area would be better able to identify likely targets for reform litigation.<sup>99</sup> In litigation and in legislatures, data on bail amounts, the types of charges lodged against defendants, defendants' income, bail amounts set, and length of pretrial detention would help to demonstrate the inequality and inefficacy of the current system.

Another inadequately documented problem in the bail context is the expansion of conditions attached to pretrial release—including the electronic monitoring of numerous defendants who are not particularly dangerous or likely to flee.<sup>100</sup> It is generally understood that the use of pretrial monitoring is expanding, in part due to decreasing costs of equipment like ankle monitors.<sup>101</sup>

---

<sup>99</sup> See *Ending the American Money Bail System*, *supra* note 36.

<sup>100</sup> See, e.g., Robert S. Gable, *Left to Their Own Devices: Should Manufacturers of Offender Monitoring Equipment Be Liable for Design Defect?*, 2009 U. ILL. J.L. TECH. & POL'Y 333, 338–39 (discussing states that require electronic monitoring in certain contexts and generally stating that electronic surveillance is “increasingly popular,” but not providing specific usage numbers); Annesley K. Schmidt, *Electronic Monitoring: What Does the Literature Tell Us?*, FED. PROB., Dec. 1998, at 10, 11 (observing that “[i]t is not known how many offenders are being monitored on a particular date, from a 1-day count, or over some time period such as a year,” and describing the incomplete studies that relied on surveys to attempt to ascertain the numbers); see also Molly Carney, Note, *Correction Through Omniscience: Electronic Monitoring and the Escalation of Crime Control*, 40 WASH. U. J.L. & POL'Y 279, 287–89 (2012) (providing examples of pilot programs and describing the increased use of monitoring but not locating or identifying statistics on the exact amount of monitoring in the pretrial or other contexts). Some data is, however, available for limited jurisdictions. BUREAU OF JUSTICE ASSISTANCE, U.S. DEP'T OF JUSTICE, OFFENDER SUPERVISION WITH ELECTRONIC TECHNOLOGY 16–17 (2d ed. 2006), [http://www.appa-net.org/eweb/docs/APPA/pubs/OSET\\_2.pdf](http://www.appa-net.org/eweb/docs/APPA/pubs/OSET_2.pdf) [<https://perma.cc/NXS2-VGF5>] (describing annual surveys by the *Journal of Offender Monitoring* that use manufacturer's reports on the number of units manufactured and deployed to “estimate” the number of offenders, not pretrial detainees, that are released and electronically supervised, and noting projections that a total of 200,000 GPS and radio frequency units would be in use in 2009); THE PEW CTR. ON THE STATES, ONE IN 31: THE LONG REACH OF AMERICAN CORRECTIONS 25 (Mar. 2009), [http://www.pewtrusts.org/~media/assets/2009/03/02/pspp\\_1in31\\_report\\_final\\_web\\_32609.pdf](http://www.pewtrusts.org/~media/assets/2009/03/02/pspp_1in31_report_final_web_32609.pdf) [<https://perma.cc/X2M3-Q85W>] (noting that electronic monitoring has become more common but only providing examples of programs that increasingly use it rather than describing numbers of offenders monitored); see also Cynthia Dizikes & Todd Lighty, *Electronic Monitoring Spikes in Cook County*, CHI. TRIB. (Feb. 23, 2015), <http://www.chicagotribune.com/news/ct-electronic-monitoring-met-20150222-story.html> [<https://perma.cc/S83S-Q2S4>] (“Judges last year ordered 14,717 men and women awaiting trial onto electronic home monitoring as a condition of their release from custody, compared with 8,657 in 2013, according to figures provided by the sheriff's office.”).

<sup>101</sup> See Erin Murphy, *Paradigms of Restraint*, 57 DUKE L.J. 1321, 1367–68 (2008) (“[T]he economics of technological control enable the regulation of greater numbers of persons under less stringent conditions for a longer period of time and to a greater degree

But empirical support to document the scope of this “net widening” problem or the jurisdictions with the largest monitoring expansions remains elusive.<sup>102</sup>

## 2. Counsel for Indigent Defendants

In the indigent defense context, where there is a long-recognized constitutional right to counsel,<sup>103</sup> bulk data could be used to identify the jurisdictions that routinely deny counsel, or provide counsel too late, to the largest percentage of indigent defendants. This would allow the targeted use of nonprofit funds in areas where the most improvement is needed, and it could potentially shame policymakers in these areas into providing additional funding for indigent defense despite the general unpopularity of this funding priority. But the existing data is, for the most part, too generalized to be useful for this purpose. Based on a 2007 survey of 1,046 public defender offices, 950 of which provided at least partial responses,<sup>104</sup> we know total expenditures on public defense by these agencies and offices that coordinate appointed counsel.<sup>105</sup> We also know the types of cases that tend to dominate county public defenders’ caseloads, among other data.<sup>106</sup> Another BJS survey-based report describes the number of states that use public defense agencies as opposed to court-appointed counsel,<sup>107</sup> the attrition rate for public defender programs,<sup>108</sup> the percentage of programs that exceeded maximum recommended annual caseloads for attorneys,<sup>109</sup> caseload increases between 1999 to 2007,<sup>110</sup> and other data. And the underlying survey data used by the BJS in these studies would point to the specific states where attorneys

---

than an equivalent physical intrusion.”); Wiseman, *supra* note 21, at 1372–73 (noting cost savings realized when counties, states, and certain federal districts increased release and monitoring and reduced pretrial detention). *But see* Eric Maes & Benjamin Mine, *Some Reflections on the Possible Introduction of Electronic Monitoring as an Alternative to Pre-Trial Detention in Belgium*, 52 HOW. J. CRIM. JUST. 144, 150–57 (2013) (arguing that electronic monitoring is costly and poses numerous privacy concerns).

<sup>102</sup> *Cf.* Wiseman, *supra* note 21, at 1368–70 (describing the limited studies on pretrial electronic monitoring and its effectiveness).

<sup>103</sup> *See* Gideon v. Wainwright, 372 U.S. 335, 344 (1963) (“[I]n our adversary system of criminal justice, any person haled into court, who is too poor to hire a lawyer, cannot be assured a fair trial unless counsel is provided for him.”).

<sup>104</sup> DONALD J. FAROLE, JR. & LYNN LANGTON, U.S. DEP’T OF JUSTICE, NCJ 231175, COUNTY-BASED AND LOCAL PUBLIC DEFENDER OFFICES, 2007, at 14 (Sept. 2010), <https://www.bjs.gov/content/pub/pdf/clpdo07.pdf> [<https://perma.cc/7FET-QP2P>].

<sup>105</sup> *Id.* at 3.

<sup>106</sup> *Id.* at 9.

<sup>107</sup> LYNN LANGTON & DONALD FAROLE, JR., U.S. DEP’T OF JUSTICE, NCJ 228229, STATE PUBLIC DEFENDER PROGRAMS, 2007, at 1 (Sept. 2010), <http://www.bjs.gov/content/pub/pdf/spdp07.pdf> [<https://perma.cc/84W3-2JXK>].

<sup>108</sup> *Id.* at 18.

<sup>109</sup> *Id.* at 12.

<sup>110</sup> *Id.* at 18.

exceeded recommended caseloads and other state-specific data.<sup>111</sup> Further, there is abundant, generalized data on the underfunding of counsel for indigent defendants,<sup>112</sup> the lack of adequate investigative support for these counsel,<sup>113</sup> and other systemic challenges. But there appears to be no available data that shows which county, state, or federal courts—or even which jurisdictions—deny counsel to the highest percentage of indigent defendants or most routinely provide counsel at too late a stage. Even where information on the appointment of counsel or lack of appointment is included within court dockets or summaries, it is often difficult to tell why counsel was not appointed, such as whether an indigency hearing was not held, the defendant waived the right to counsel, or the defendant was determined to have adequate funds.

The groups that have succeeded in showing routine denials of counsel in certain jurisdictions, such as within the Williamson County and Gillespie County, Texas courthouses, have had to physically send their attorneys to courts to monitor court proceedings.<sup>114</sup> This type of effort may still be necessary, but if the types of data described were readily available, its analysis would direct and support policy reforms and litigation.

### 3. Debtors' Prisons

There is a clear federal<sup>115</sup> and state constitutional right<sup>116</sup> against the jailing of defendants who cannot afford to pay fees related to criminal matters on the basis of failure to pay alone. Yet anecdotal evidence and some limited statistical analyses show that imprisoning defendants who fail to pay probation fees and fines (often levied by private probation agencies), child support, and similar debts is an increasingly common practice.<sup>117</sup> Several organizations have begun conducting broad analyses of the jurisdictions that appear to most commonly jail defendants for a failure to pay, with the American Civil

---

<sup>111</sup> *Id.* at 12.

<sup>112</sup> *See, e.g.*, STANDING COMM. ON LEGAL AID & INDIGENT DEFENDANTS, AM. BAR ASS'N, *GIDEON'S BROKEN PROMISE: AMERICA'S CONTINUING QUEST FOR EQUAL JUSTICE* 9–10 (Dec. 2004), [http://www.americanbar.org/content/dam/aba/administrative/legal\\_aid\\_indigent\\_defendants/ls\\_sclaid\\_def\\_bp\\_right\\_to\\_counsel\\_in\\_criminal\\_proceedings.authcheckdam.pdf](http://www.americanbar.org/content/dam/aba/administrative/legal_aid_indigent_defendants/ls_sclaid_def_bp_right_to_counsel_in_criminal_proceedings.authcheckdam.pdf) [<https://perma.cc/V9J5-VMWJ>]; *see also* NAT'L RIGHT TO COUNSEL COMM., *supra* note 21, at 99–100.

<sup>113</sup> LANGTON & FAROLE, *supra* note 107, at 1.

<sup>114</sup> *See supra* notes 25, 37 and accompanying text. The Texas Fair Defense Project, which filed the cases exposing these access to counsel problems, was able to identify the lack of appointment of counsel, but could not ascertain the reasons for the failure to appoint counsel, whether it was because a defendant waived counsel or hired private counsel. E-mail from Harry Williams, to author (Feb. 19, 2016) (on file with author).

<sup>115</sup> *Bearden v. Georgia*, 461 U.S. 660, 672–73 (1983).

<sup>116</sup> *See Wagner, supra* note 21, at 383 (“Today most state constitutions explicitly forbid imprisonment for civil debts.”).

<sup>117</sup> *See infra* notes 118–23 and accompanying text.

Liberties Union (ACLU) investigating Georgia, Louisiana, Michigan, Ohio, and Washington State.<sup>118</sup> The organization collected its data based on a survey of “state ACLU affiliates asking them to provide any information they had on debtors’ prisons in their state,” and it posted a similar survey on a National Legal Aid and Defender Association listserv.<sup>119</sup> The ACLU selected these states based on the survey responses it received and then, from December 2009 through July 2010, relied upon “ACLU affiliate staff, law students, volunteer attorneys, and law professors” to investigate practices in the identified states through review of local and state case dockets and pleadings, and interviews of individuals involved in these states’ criminal justice systems.<sup>120</sup> Thus, approximately a year and a half of heavy legwork resulted in case studies and descriptions of certain debtors’ prison problems in five states. The ACLU subsequently filed lawsuits challenging debtors’ prison in several jurisdictions, such as Biloxi, Mississippi;<sup>121</sup> and Benton County, Washington.<sup>122</sup>

The Southern Center for Human Rights has similarly sued private probation companies, diversion centers, and cities in an effort to stop the practice of imprisoning defendants for a failure to pay criminal fees and fines.<sup>123</sup> Although on the ground investigation will always be necessary, bulk data would make the jurisdictions with the most problematic practices easier to identify and would allow scarce resources to be put to the best use. Electronic information on the amount of probation fine or fee, child support, or other debt owed; the result (imprisonment or otherwise); the length of the sentence; and other data could provide a more accurate, detailed, and comprehensive picture of the problem with much less effort.

Analysis of better data could also help identify new, localized problems in the criminal justice area that are as-yet unnoticed or underexplored. For example, although the DOJ often researches and suggests widespread, national reforms, it also targets certain jurisdictions for specific reforms. The DOJ began its investigation of Ferguson, Missouri’s juvenile justice system only after the highly publicized shooting of Michael Brown by a police officer.<sup>124</sup>

---

<sup>118</sup> AM. CIVIL LIBERTIES UNION, *supra* note 36, at 5.

<sup>119</sup> *Id.* at 13.

<sup>120</sup> *Id.*

<sup>121</sup> See generally Class Action Complaint, *Kennedy v. City of Biloxi*, No. 1:15-cv-348-HSO-JCG (S.D. Miss. Oct. 21, 2015).

<sup>122</sup> See generally Complaint for Declaratory and Injunctive Relief, *Fuentes v. Benton County*, No. 15-2-02976-1 (Wash. Super. Ct. Oct. 7, 2015).

<sup>123</sup> Complaint at 2, *Edwards v. Red Hills Cmty. Prob., LLC*, No. 1:15-cv-00067-LJA (M.D. Ga. Apr. 10, 2015); Amended Complaint at 2, *Thomas v. City of Gulfport*, No. 1:05-CV-349-LG-RHW (S.D. Miss. Aug. 9, 2005); Petition for a Writ of Habeas Corpus at 3, *Hurley v. Hinton*, No. HC00532 (Ga. Super. Ct. Sept. 2006), [https://www.schr.org/files/hurley\\_habeas.pdf](https://www.schr.org/files/hurley_habeas.pdf) [<https://perma.cc/27BR-EXPW>].

<sup>124</sup> See Matt Apuzzo, *Department of Justice Sues Ferguson, Which Reversed Course on Agreement*, N.Y. TIMES (Feb. 10, 2016), <https://www.nytimes.com/2016/02/11/us/politi>

The DOJ reviewed 32,169 cases (and more than 14,000 pages of documents),<sup>125</sup> interviewed court personnel, attended a juvenile detention hearing, read transcripts of seventy proceedings, met with state and local public defender staff, spoke with representatives of juvenile law clinics, interviewed parents of several youth who had gone through the family court system, and reviewed existing Missouri-specific reports to identify a variety of problems.<sup>126</sup> These include, for example, disproportionate representation of African American juvenile defendants in pretrial detention and several other “major decision points” in juvenile justice, such as post-adjudication placement of defendants in youth services custody.<sup>127</sup> The DOJ also observed that indigent youth in Missouri routinely lack access to counsel at critical stages of their cases, noting that “[a]pproximately 60% of lawyers who represent detained youth in Family Court enter the case at least seven days – and as many as 232 days – after the child’s detention hearing is held.”<sup>128</sup> Many jurisdictions that have not received attention focused on Ferguson likely suffer from similar problems—problems that could be at least partially revealed by readily available bulk criminal justice data—that remain undetected.

### C. Bulk Data and Reform

If docket-level data on criminal justice cases were collected in a uniform manner by even just a few states and large county courts and made electronically available, this would provide scholars, policymakers, and reform groups with a clearer picture of the criminal justice system without requiring time-intensive surveys, visits to courtrooms, interviews, and the other methods commonly relied on now. For local reform, on the ground investigation would likely remain essential, but bulk data analysis could provide the initial picture needed to point investigations toward the most problematic areas. Further, where investigative and anecdotal data already exists, as it does for bail, indigent counsel, and debtors’ prisons, empirical data would provide important support for reform arguments and legal complaints that currently must rely primarily on the anecdotal data.

Outside the litigation context, extensive criminal justice data could also trigger legislators to introduce reform legislation and provide needed evidence for legislative reforms. Reform advocates currently have to rely on summary statistics to support their arguments, such as the often-cited statistic that two-

---

cs/justice-department-sues-ferguson-over-police-deal.html?\_r=0 [https://perma.cc/D7MH-V2H7] (“The shooting was deemed justified, but it . . . brought the attention of the Department of Justice to the city’s police practices.”); see also CIVIL RIGHTS DIV., *supra* note 28, at 5 n.6.

<sup>125</sup> CIVIL RIGHTS DIV., *supra* note 28, at 12, 34.

<sup>126</sup> *Id.* at 12–13.

<sup>127</sup> *Id.* at 35.

<sup>128</sup> *Id.* at 15.

thirds of the individuals in jail at any given time are individuals who are awaiting trial but are low-risk defendants.<sup>129</sup> More specific data—particularly data highlighting high pretrial detention rates in certain jurisdictions despite defendants in that jurisdiction not being more likely to flee or be violent—could help to support state and county legislative reforms.

Finally, the data could directly change the behavior of police, hearing officers, attorneys, and judges, particularly the many judges at the state and county level that are elected. With knowledge that statistics will be systematically collected, disclosed, and analyzed, officials in the criminal justice system might have a greater incentive to avoid constitutional violations.<sup>130</sup> Further, individuals within the criminal justice system who want to produce better results but are underresourced and overburdened could use big data to identify targeted improvements. As Pamela Metzger and Andrew Ferguson note, public defenders could benefit from systemically collected data that might show the areas in which avoidable errors tend to occur most frequently,<sup>131</sup> and the types of case strategies—such as spending time interviewing witnesses versus filing more motions that are more likely to lead to positive outcomes—and thus should potentially be the focus of attorneys' limited resources.<sup>132</sup> Professor Ferguson similarly believes that the use of big data could improve jury selection, making juries more representative of the community, among other improvements.<sup>133</sup> The influence of statistics on police, attorneys, juries, judges, and other individuals within the criminal justice system will likely not be as powerful as more direct signals, such as body cameras on police<sup>134</sup> or the docking of pay for attorneys who make repeated errors, but it should have some deterrent power.

In order to capture these benefits, however, there must be significant changes in how criminal justice data is collected, stored, and made public because the available data does not allow for the type of analysis that would guide reform priorities and further inform litigation and policy-based efforts.

---

<sup>129</sup>CRIMINAL JUSTICE SECTION, *supra* note 27, at 2.

<sup>130</sup>Some believe that excluding evidence collected in violation of the Fourth Amendment has only mild effects on police behavior. *See, e.g.*, *United States v. Leon*, 468 U.S. 897, 918 (1984); *United States v. Janis*, 428 U.S. 433, 454 (1976); William C. Heffernan & Richard W. Lovely, *Evaluating the Fourth Amendment Exclusionary Rule: The Problem of Police Compliance with the Law*, 24 U. MICH. J.L. REFORM 311, 358–59 (1991). But more direct data pointing to constitutional violations committed by specific officers, and widely publicized data, might have a slightly greater deterrent effect.

<sup>131</sup>Metzger & Ferguson, *supra* note 42, at 1099–100.

<sup>132</sup>*Id.* at 1097–118.

<sup>133</sup>Andrew Guthrie Ferguson, *The Big Data Jury*, 91 NOTRE DAME L. REV. 935, 935–36 (2016).

<sup>134</sup>*See, e.g.* PRESIDENT'S TASK FORCE ON 21ST CENTURY POLICING, *supra* note 32, at 31–32 (noting that a “randomized control trial” “found that the officers wearing the cameras had 87.5 percent fewer incidents of use of force and 59 percent fewer complaints than the officers not wearing the cameras”).

The following Part describes the limited data that is currently collected in the criminal justice areas in which it could offer the most value.

### III. EXISTING CRIMINAL JUSTICE DATA SOURCES

As described in the preceding Part, there are significant potential benefits to collecting and disseminating basic, uniform information data from the local, state, and federal levels.<sup>135</sup> Based on the most pressing issues addressed above, some of the most valuable data would include defendant demographics such as age, race, income level, family ties, and job; type and date of charges; date of appointment of counsel, if counsel was appointed, and reasons for a failure to appoint counsel; date of the bail hearing; conditions of pretrial release and the amount of bail set or pretrial detention and length of detention; whether or not the defendant appeared for trial or committed pretrial crimes; plea bargain or conviction date; sentence and length of sentence; date of release or release on parole; and parole conditions and outcomes.<sup>136</sup> As discussed in this Part, many courts collect some of this data within dockets, and jails and prisons retain certain statistics about the defendants within their systems, but they do not routinely collect all of this data or record it in the same way. In part due to the lack of uniform data collection, and in part due to the failure of those who collect this data to make it electronically available, the existing data is difficult, if not impossible, to systematically acquire or analyze.

#### A. *Data Sources*

Several groups already report or collect and synthesize criminal justice data, but not in consistent ways, and often without making it available electronically. State and local courts and court administrators already produce some of the data most important for the criminal justice reform efforts identified in Part I. With some modification, including more uniformity with respect to what is reported and, critically, in the manner of disseminating it, this information could be quite valuable to criminal justice study and reform. Although commercial entities already collect and compile massive databases of criminal justice information, they tend to be too expensive for bulk searches, and to focus only on information needed for background checks, such as convictions.

#### 1. *State and Local Governments*

Data collection and distribution policies vary substantially across state and local governments. In general, the institutions that routinely interact with criminal defendants already collect and record some or all of the necessary

---

<sup>135</sup> See *supra* Part II.

<sup>136</sup> *Id.*



information. In addition to collecting data, courts at all levels have historically provided public access to records.<sup>137</sup> This is due in part to Supreme Court doctrine, which provides a “right to inspect and copy public records and documents, including judicial records and documents,” while allowing each court to deny access where court files might be used for “improper purposes.”<sup>138</sup> Courts have allowed public access for reasons of tradition, public trust, and integrity, among others.<sup>139</sup> But many courts, police departments, jails, and prisons collect only a portion of the data necessary to provide a relatively comprehensive picture of problems in the criminal justice system.<sup>140</sup> Moreover, the data is collected and reported in nonuniform ways, making large-scale analysis more difficult. And even if the records exist, access is often a problem. Some local and state courts provide no online records;<sup>141</sup> others collect limited information about defendants and their cases but provide only limited electronic access for a fee and prevent or limit bulk access<sup>142</sup> in order to avoid commercial misuse of defendants’ private information.<sup>143</sup> But the barriers are much higher than they need to be,

---

<sup>137</sup>Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137, 1145 (2002).

<sup>138</sup>*Nixon v. Warner Commc’ns, Inc.*, 435 U.S. 589, 597–98 (1978) (footnote omitted). For a thorough examination of public access to court records and limitations on this access as a result of Supreme Court doctrine and the common law, see Solove, *supra* note 137, at 1154–60.

<sup>139</sup>See MD. COURTS, REPORT OF THE COMMITTEE ON ACCESS TO COURT RECORDS 6, <http://www.courts.state.md.us/access/finalreport3-02.pdf> [<https://perma.cc/LJ76-PWLP>].

<sup>140</sup>See *infra* notes 147–48.

<sup>141</sup>See, e.g., *Privacy/Public Access to Court Records*, NAT’L CTR. FOR ST. CTS., <http://www.ncsc.org/Topics/Access-and-Fairness/Privacy-Public-Access-to-Court-Records/State-Links.aspx?cat=Public%20Access%20Web%20Sites#New%20Hampshire> [<https://perma.cc/4SAN-U4GQ>] (concluding that Maine, New Hampshire, and North Carolina have “no online court records”).

<sup>142</sup>See, e.g., ARIZ. CODE OF JUDICIAL ADMIN. pt. I, ch. 6, § 1-605(D) (2010), [http://www.azcourts.gov/Portals/0/admcode/pdfcurrentcode/1-605\\_New\\_Dec\\_09.pdf](http://www.azcourts.gov/Portals/0/admcode/pdfcurrentcode/1-605_New_Dec_09.pdf) [<https://perma.cc/Y65G-3D2L>] (requiring a “dissemination agreement” for requests for bulk data between the requestor and the bulk data custodian, as well as a “subscription service”); CAL. R. CT. 2.503(f)–(g) (“The court may only grant electronic access to an electronic record . . . on a case-by-case basis. . . . The court may provide bulk distribution of only its electronic records of a calendar, register of actions, and index.”); N.H. CT. R. GUIDELINES FOR PUBLIC ACCESS TO COURT RECORDS, <http://www.courts.state.nh.us/rules/misc/misc-8.htm> [<https://perma.cc/DKM5-54SQ>] (“Access to large numbers of records at any one time shall not be permitted.”).

<sup>143</sup>See, e.g., MD. COURTS, *supra* note 139, at 1 (“These technological advances have led to increased concerns about, among other things, protection of privacy, protection against risk of harm, and the consequences of sale by businesses of information from court records.”); SEARCH, THE NAT’L CONSORTIUM FOR JUSTICE INFO. & STATISTICS, REPORT OF THE NATIONAL TASK FORCE ON THE COMMERCIAL SALE OF CRIMINAL JUSTICE RECORD INFORMATION 46 (2005) [hereinafter SEARCH], <http://www.search.org/files/pdf/RNTFCS>

particularly in light of the fact that private information could simply be omitted from the data prior to electronic posting.

The records most likely to contain the information needed for big data analysis in the criminal justice system are case summaries and dockets.<sup>144</sup> Many contain information such as the charge and degree of felony charged; and the statute under which the defendant was charged; case dispositions and pleas; the maximum prison sentence and length of confinement; years of probation; penalties and conditions imposed such as community service and drug treatment and randomized testing; subsequent events such as violation of probation.<sup>145</sup> Court summaries from county and state courts also sometimes include the date on which a surety bond was posted, if any; dates of request for counsel for an indigent defendant; dates of filing of plea forms; and dates on which the judgment and sentence were issued and commitment occurred, among other data.<sup>146</sup> These records do not, however, consistently<sup>147</sup> include important data that show when and if counsel was appointed for indigent defendants, the amount of the bond set by the judge (if bond was set),<sup>148</sup> and

CJRI.pdf [<https://perma.cc/TUJ6-PVP6>] (“In attempting to balance the competing interests posed by new technologies, privacy interests, and historical principles of openness, the courts have grappled with several questions.”).

<sup>144</sup> SEARCH, *supra* note 143, at 45 (“[T]he primary source of criminal history information for commercial vendors is the court system.”).

<sup>145</sup> See, e.g., *Case Summary: Case No. 89-CF-009822-A*, HILLSBOROUGH COUNTY CTS. REC. SEARCH, <http://pubrec10.hillsclerk.com/Unsecured/CaseDetail.aspx?CaseID=4052837> [<https://perma.cc/3NR2-8NYD>] (showing court appointed counsel, sentence, and case disposition); see also *infra* notes 147–48.

<sup>146</sup> See, e.g., *Case Summary: Case No. 88-CF-004404-A*, HILLSBOROUGH COUNTY CTS. REC. SEARCH, <http://pubrec10.hillsclerk.com/Unsecured/CaseDetail.aspx?CaseID=7848687> [<https://perma.cc/PA58-GM77>].

<sup>147</sup> Some, however, do include this data. See, e.g., *Case Details – Summary: Case # 2016NY040846*, N.Y. ST. UNIFIED CT. SYS., [https://iapps.courts.state.ny.us/webcrim\\_attorney/Detail?which=case&docketNumber=JvwWNX464YN2B8Fd/2okRQ==&countyId=ftUCLXxA/VgUxT4CDSVwBw==&docketId=RGepsZhfKEAkPaKV8x\\_PLUS\\_DRA==&docketDseq=o6PDyKIx4BvSfbvyCgDnHw==&defendantName=Johnson,+Joseph+E&court=New+York+Criminal+Court&courtType=L&recordType=C&recordNum=4YsLz49JROPZoddX8PjTVw==](https://iapps.courts.state.ny.us/webcrim_attorney/Detail?which=case&docketNumber=JvwWNX464YN2B8Fd/2okRQ==&countyId=ftUCLXxA/VgUxT4CDSVwBw==&docketId=RGepsZhfKEAkPaKV8x_PLUS_DRA==&docketDseq=o6PDyKIx4BvSfbvyCgDnHw==&defendantName=Johnson,+Joseph+E&court=New+York+Criminal+Court&courtType=L&recordType=C&recordNum=4YsLz49JROPZoddX8PjTVw==) [<https://perma.cc/YCM5-E62F>] (showing the type of attorney appointed, such as “Legal Aid,” although not showing the date on which the attorney was assigned to the defendant’s case).

<sup>148</sup> *But see*, e.g., *Charleston County Circuit Court Case Details Public Index: Case Number 2016A1010200127*, CHARLESTON COUNTY CLERK CT., <http://jemsweb.charlestoncounty.org/publicindex/> (click “Accept,” then enter case number 2016A1010200127) [<https://perma.cc/5XXQ-YJ3N>] (providing data on bond type and amount); *Docket No. D03D-CR16-0154213-S*, ST. CONN. JUD. BRANCH, <http://www.jud2.ct.gov/crdockets/DocketNoEntry.aspx?source=Pend> (enter docket number D03D-CR16-0154213-S to search) [<https://perma.cc/SJ2Z-CXXE>] (providing individual pretrial case data that indicates defendant’s birth year, arrest date, bond amount and type, charges and statute under which the defendant was charged, and the class and type of the charge, e.g., misdemeanor, although the information must be located using a docket number or defendant’s name and is only provided on a case-by-case basis).

information necessary for identifying problems in the system such as pretrial detention and unaffordable bail amounts for low-income defendants.

As introduced above, beyond differences in information collection, much of the case summary and docket data available from states and local governments is not easily accessible online.<sup>149</sup> When county and state databases with criminal case information do exist, they typically require individuals searching for a case summary to input the name of the defendant or the case number.<sup>150</sup> Access to the full database, which could be downloaded

---

<sup>149</sup> See, e.g., *supra* note 141 and accompanying text.

<sup>150</sup> See, e.g., *Appellate Courts: General Docket*, ST. MISS. JUDICIARY, [http://courts.ms.gov/appellate\\_courts/generaldocket.html](http://courts.ms.gov/appellate_courts/generaldocket.html) [<https://perma.cc/9UGB-EWN7>] (requiring case year and sequence number, party name, or attorney name for a search); *Case Lookup*, N.M. CTS., <https://caselookup.nmcourts.gov/caselookup/app> [<https://perma.cc/7FZY-MBJ8>] (providing only “criminal Domestic Violence and DWI historic convictions from September 1, 1991 onwards” and later for municipal courts, and requiring a search by name or case number); *Case Search*, CLERK SUPERIOR ST. & MAGISTRATE CTS. COUNTY GWINNETT, <http://www.gwinnettcourts.com/casesearch/> [<https://perma.cc/9GMC-MZGC>] (allowing individual searches by case number, party name, attorney, judge, or scheduled events); *Case Search*, MICH. CTS., [http://courts.mi.gov/opinions\\_orders/case\\_search/pages/default.aspx](http://courts.mi.gov/opinions_orders/case_search/pages/default.aspx) [<https://perma.cc/UDT5-2N87>] (requiring a search by docket number or party name); *Case Search*, MYCASE.IN.GOV, <https://mycase.in.gov/Search.aspx?ID=100> [<https://perma.cc/R7RD-DF72>] (providing access to certain Indiana courts but requiring an individual case number, a party, or attorney name for a search); *Circuit Courts of Illinois*, JUDICI, [https://www.judici.com/courts/illinois\\_courts.jsp](https://www.judici.com/courts/illinois_courts.jsp) [<https://perma.cc/Y7KM-733U>] (last modified Apr. 17, 2017) (providing limited free data for approximately eighty Illinois circuit courts but requiring payment for “multi-court/multi-case results”); *CourtConnect*, ARK. JUDICIARY, <https://courts.arkansas.gov/administration/acap/courtconnect> [<https://perma.cc/5XFR-LLDN>] (providing full information from some county courts and partial information from others, but requiring a search by case name or number); *Court Case Search*, WELCOME TO CHANDLER, <http://www.chandleraz.gov/cjis/courtcasesearch/> [<https://perma.cc/VZU6-42VF>] (requiring searches by case number or defendant name for cases in Chandler, Arizona); *Criminal Case Summary*, SUPERIOR CT. CAL., <http://www.lacourt.org/criminalcasesummary/ui/index.aspx> [<https://perma.cc/RB2P-C4RT>] (requiring the case number); *Criminal Cases: PROMIS/Gavel Public Access: Search Menu*, N.J. CTS. PUB. ACCESS, <https://mccs.njcourts.gov/webe4/ExternalPGPA/> [<https://perma.cc/F9QD-4A7J>] (providing access only to criminal cases for defendants indicted and convicted or sentenced in New Jersey Superior Court, and requiring a search by name, indictment/accusation number, or court disposition report (CDR) number); *Criminal/Motor Vehicle Case Look-Up*, ST. CONN. JUD. BRANCH, <http://www.jud.ct.gov/crim.htm> [<https://perma.cc/Q9VJ-89QK>] (requiring search by defendant or docket number for convictions and pending cases); *Idaho Supreme Court Data Repository*, *supra* note 46 (providing case information for all Idaho courts with the exception of Twin Falls and Ada County, searchable individually by name or case number, and maintaining confidentiality by omitting “the first six characters of social security numbers, street addresses, telephone numbers, [and] any personal identification numbers,” such as “motor vehicle operator’s license numbers”); *Iowa Courts Online Search*, IOWA CTS., <https://www.iowacourts.state.ia>

and searched for all cases, typically requires permission of the court administrator.<sup>151</sup> Many state and local courts also charge a fee for access to individual criminal case summaries.<sup>152</sup>

---

.us/ESAWebApp/TrialSimpFrame [https://perma.cc/B976-2BPS] (providing case information for all Iowa courts but requiring search by name, case ID, or citation number); *Maryland Judiciary Case Search Criteria*, MD. JUDICIARY, <http://casesearch.courts.state.md.us/casesearch/inquirySearch.jis> [https://perma.cc/VW7R-YGFV] (requiring entry of a party name or case number in order to search circuit or district court records); *Minnesota Trial Court Public Access (MPA) Remote View*, MINN. JUD. BRANCH, <http://pa.courts.state.mn.us/Search.aspx?ID=100> [https://perma.cc/J5Y3-P67H] (requiring case number, defendant name, citation number, or attorney name for a search); *Online Court Records Search*, CIVITEKFLORIDA.COM, <https://www.civitekflorida.com/ocrs/county/index.xhtml> [https://perma.cc/M9RJ-ZJEL] (providing online court data for courts in approximately thirty-two Florida counties); *Online Court Records Search*, FRANKLINCLERK.COM, <https://www.civitekflorida.com/ocrs/app/search.xhtml> [https://perma.cc/TP7N-MYDL] (requiring search by the defendant's first and last name or the case sequence number for cases in Franklin County, Florida); ON DEMAND CT. RECS., <http://www1.odcr.com/> [https://perma.cc/45H8-NMGP] (requiring party name, case number, or date for searches of court records for cases in Oklahoma); *Public Access to Court Information*, ARIZ. JUD. BRANCH, <https://apps.supremecourt.az.gov/publicaccess/%28X%281%29S%28x013i1m03tcgts55sx32hci3%29%29/caselookup.aspx> [https://perma.cc/9VMK-ZVAG] (offering access to court records from all but twelve county and state courts in Arizona but requiring searches by name or case number); *Public Case Information*, SUPREME JUD. CT. & APPEALS CT. MASS., <http://www.mappellatecourts.org/index.php> [https://perma.cc/ZM7U-536R] (requiring a search by involved party or docket number); *Public View Docket*, ST. MONT. OFF. CLERK SUP. CT., <https://supremecourtdocket.mt.gov/activecase.jsp> [https://perma.cc/M4VD-AJ84] (requiring case number, party name, or attorney name); *Search Cases*, ST. LA., <http://www.la4th.org/SearchCases.aspx> [https://perma.cc/N42W-U7DW] (requiring a search by district or appellate court case number or attorney bar roll number); *Smart Search*, R.I. JUDICIARY PUB. PORTAL, <https://publicportal.courts.ri.gov/PublicPortal/Home/Dashboard/29> [https://perma.cc/7LQL-BRFW] (requiring search by party name or record number); *WebCriminal*, N.Y. ST. UNIFIED CT. SYS., <http://manhattanda.org/case-information> (requiring a case number or defendant name) [https://perma.cc/KKP2-QFQ5].

<sup>151</sup> *CourtView: Justice Solutions*, FAIRFIELD COUNTY CLERK CTS., <http://courtview.co.fairfield.oh.us/eservices/search.page.3?x=nojSzI8grhRfsG5ne3CyIw> [https://perma.cc/2YDR-BNB7] (allowing searches by case type); *Detailed Case Search*, SECOND JUD. DISTRICT CT. ST. NEV., <https://www.washoecourts.com/index.cfm?page=caserearch&CaseNum=&CaseCapt=&CaseType=CR> [https://perma.cc/DJ66-WDXW] (allowing a broad case search by "case type," such as "criminal," which produces numerous cases and associated data); *Public Index*, CHARLESTON COUNTY, <http://jcmsweb.charlestoncounty.org/PublicIndex/PISearch.aspx> [https://perma.cc/DZE4-YBNA] (allowing searches by date range, which produces numerous criminal dockets); *Your Missouri Courts*, MO. CASE.NET, <https://www.courts.mo.gov/casenet/base/welcome.do> [https://perma.cc/VSW2-YAS9] (allowing for a search by filing date, which produces numerous cases and data about each case).

<sup>152</sup> See, e.g., *District Court Records Search*, KAN. OFF. JUD. ADMIN., <https://www.kansas.gov/countyCourts/> [https://perma.cc/U3ZK-UMNM] ("The cost is \$1.00 per search and \$1.00 per case retrieved for view."); *Just One Look*, ALACOURT.COM, <https://pa.alacourt.com/Default.aspx> [https://perma.cc/TLA8-SQ4U] (indicating a cost of

Some states require uniform reporting and data collection by all courts within the state—a practice that, if followed in all states, would substantially improve data availability. For example, through New York State’s Unified Court System, “[c]ity, district, county, supreme courts, courts acting as appellate courts, and the Court of Appeals have a responsibility to report the criminal case outcome to the Unified Court System which maintains a repository of court case information.”<sup>153</sup> Certain courts “report information directly to the state repository,” whereas others send data to the Office of Court Administration (OCA); this office then sends the data in electronic form to the state repository.<sup>154</sup> For each defendant in the system with arrest information, the court action closing out that defendant’s case also must be included in the system.<sup>155</sup> Specific information collected and reported by all local and state courts in New York to the repository includes “Defendant Pedigree Information,” which consists of “[n]ame and [a]lias(es),” sex, race, and “[a]ge or date of birth, if available and verifiable.”<sup>156</sup> The information also must include “Criminal History Identifiers” including a unique record number that the state assigns to each defendant,<sup>157</sup> other tracking numbers, and the arrest date. The repository contains arrest information such as date and place of the crime and “[d]etailed [a]rrest [c]harge information,” as well as information about whether the arrestee is a juvenile.<sup>158</sup> And finally, the court disposition information in the repository includes the name of the court and detailed “[a]rrangement or [d]isposition [c]harge information” including the title, section, and subsection of the state law under which the defendant was charged, the class and category of charge, the degree of the charge, and the number of counts, among other data points.<sup>159</sup>

New York’s repository lacks some useful data, such as pretrial release or detention information. Nonetheless, as a compilation of uniform records from all courts within the state, it is a model for improved information collection

---

\$9.99 for a name or case number search in Alabama court records); *Nebraska Trial Courts Online Case Search*, NEB. JUD. BRANCH, <https://www.nebraska.gov/justicecc/ccname.cgi> [<https://perma.cc/KB4Z-GREE>] (providing information on up to thirty cases for a \$15 search fee); *OJIN OnLine*, OR. CTS., <http://courts.oregon.gov/OJD/OnlineServices/OJIN/pages/index.aspx> [<https://perma.cc/X6LC-YFWS>] (granting access to criminal and civil records for a subscription); *Public Case Access System*, SACRAMENTO SUPERIOR CT., <https://services.saccourt.ca.gov/PublicCaseAccess/> [<https://perma.cc/4VJ5-8QMR>] (noting public access fees).

<sup>153</sup> *Division of Criminal Justice Services: Introduction*, N.Y. ST., <http://www.criminaljustice.ny.gov/stdpractices/jud.htm> [<https://perma.cc/F8F8-H3WD>].

<sup>154</sup> *Id.*

<sup>155</sup> *Id.*

<sup>156</sup> *Id.*

<sup>157</sup> *Id.*; see also LEGAL ACTION CTR., YOUR NEW YORK STATE RAP SHEET 11 (2015), [http://lac.org/wp-content/uploads/2014/12/Your\\_New\\_York\\_State\\_Rap\\_Sheet\\_2013.pdf](http://lac.org/wp-content/uploads/2014/12/Your_New_York_State_Rap_Sheet_2013.pdf) [<https://perma.cc/D7TM-7KS2>].

<sup>158</sup> *Division of Criminal Justice Services: Introduction*, *supra* note 153.

<sup>159</sup> *Id.*

efforts. The state has important measures in place to ensure that inputs are accurate and reported using the same terms (for example, the same descriptions of the type of sentence imposed), and is corrected and updated when better information about a particular case becomes available.<sup>160</sup> To help achieve these goals, the state provides basic instructions to the courts and police on reporting requirements<sup>161</sup> and information about how to make corrections.<sup>162</sup> The computer system run by the OCA—which collects records from certain courts and provides it to the centralized state repository—also produces automated warnings when inputs appear to have anomalies, such as “conflicting information or unmatched cases.”<sup>163</sup> Further, the state directs the OCA and the Department of Criminal Justice Services, which maintains the statewide repository of criminal justice data, to operate an “ongoing reconciliation process” to ensure uniform reporting and that there are not discrepancies between the two systems.<sup>164</sup>

Unfortunately, however, New York’s criminal justice records are not readily electronically accessible to researchers and reformers. As is the case in other states, bulk searches that would unearth numerous individual criminal case summaries are not permitted: the individual defendant’s name must be known and entered for each summary to be located.<sup>165</sup> While this practice helps to protect the privacy of individual defendants, many aspects of the repository could be publicly available—with identifying information removed—without compromising defendants’ privacy. Similarly, Pennsylvania, which also has a unified court information system, permits only individual searches of criminal docket sheets by specific case identifying information, such as docket number or participant name.<sup>166</sup>

Some local governments provide similar types of information online,<sup>167</sup> although these systems are generally subject to the limitations described

---

<sup>160</sup> *Division of Criminal Justice Services: Judicial Processing – Reporting to UCS and DCJS – Monitoring Criminal Cases*, N.Y. ST., <http://www.criminaljustice.ny.gov/stdpractices/jud2a.htm> [<https://perma.cc/7BLZ-NEHJ>].

<sup>161</sup> *Division of Criminal Justice Services: Judicial Processing – Criminal Case Documentation*, N.Y. ST., <http://www.criminaljustice.ny.gov/stdpractices/jud1a.htm> [<https://perma.cc/2XDD-NE89>]; *Division of Criminal Justice Services: Judicial Processing – Reporting to UCS and DCJS – General Reporting Requirements*, N.Y. ST., <http://www.criminaljustice.ny.gov/stdpractices/jud2c.htm> [<https://perma.cc/W4AF-GBUY>]; *Division of Criminal Justice Services: Judicial Processing – Reporting to UCS and DCJS – Reporting the Sentence Imposed*, N.Y. ST., <http://www.criminaljustice.ny.gov/stdpractices/jud2g.htm> [<https://perma.cc/T7PR-NW9F>].

<sup>162</sup> *Division of Criminal Justice Services: Judicial Processing – Reporting to UCS and DCJS – Monitoring Criminal Cases*, *supra* note 160.

<sup>163</sup> *Id.*

<sup>164</sup> *Id.*

<sup>165</sup> *WebCriminal*, *supra* note 150.

<sup>166</sup> *Common Pleas Courts Docket Sheets*, UNIFIED JUD. SYS. PA., <https://ujportal.pacourts.us/DocketSheets/CP.aspx> [<https://perma.cc/M5VZ-WYKG>].

<sup>167</sup> *See supra* notes 145–52.

above. Even if bulk data from individual localities could be easily obtained, cross-jurisdiction analysis would be difficult in light of the lack of uniform collection and reporting. Nonetheless, these local systems demonstrate the feasibility of collecting and disseminating criminal justice data and could be adapted for use in the expanded, uniform reporting system proposed below.

An additional, somewhat more limited source of criminal justice information, potentially relevant to identifying incarceration for poverty, is state corrections agencies. Their records often include the type of offense with which each inmate has been charged—sometimes listing only the most serious offense as well as “the release date and whether the individual was released on parole.”<sup>168</sup> In some cases these records are only available for inmates currently held,<sup>169</sup> and are generally not electronically accessible in bulk.<sup>170</sup>

Finally, all states maintain criminal history repositories, but these often only include data on felonies and serious misdemeanors<sup>171</sup> and are more helpful to background searches than to the type of criminal justice research that is the focus of this Article.

## 2. *The Federal Government*

Some county- and state-specific criminal justice data is available as a result of data collection by federal entities. For example, the federal government collects certain consistent, uniform data on criminal cases through the BJS’s State Court Processing Statistics Program,<sup>172</sup> which

tracks felony defendants from charging by the prosecutor until disposition of their cases (a maximum of 12 months for nonmurder cases and 24 months for murder cases). Data are obtained on demographic characteristics, arrest charges, criminal justice status at time of arrest, prior arrests and convictions, pretrial release and detention, court appearance record, rearrests while on pretrial release, type and outcome of adjudication, and type and length of sentence.<sup>173</sup>

But this data is collected only for the jurisdictions representing forty of the seventy-five largest counties, and only includes felonies—“[a]ny offense punishable by death or imprisonment for more than one year”—and thus misses many of the more minor offenses.<sup>174</sup> As Professor Alexandra Natapoff

---

<sup>168</sup> SEARCH, *supra* note 143, at 25.

<sup>169</sup> *Id.*

<sup>170</sup> *Id.* at 49.

<sup>171</sup> *Id.* at 25–26.

<sup>172</sup> *Data Collection: State Court Processing Statistics (SCPS)*, *supra* note 15.

<sup>173</sup> *Id.*

<sup>174</sup> *Felony Defendants*, BUREAU JUST. STAT., <http://www.bjs.gov/index.cfm?ty=tp&tid=231> [<https://perma.cc/T7WJ-75MZ>].

has shown, these minor offenses can have serious collateral consequences.<sup>175</sup> Moreover, they can be a source of routine constitutional violations. Many individuals in jail are defendants held pretrial for nonviolent, nonfelony offenses.<sup>176</sup> Examples of defendants charged with petty crimes, such as theft of goods from stores, and unable to pay small bail amounts, fines, and fees abound. National Public Radio documents the case of a man who tried to steal a television from Walmart while high on methamphetamine.<sup>177</sup> He was unable to pay the \$150 in bail set for his case, and, at the time of the story, had been in jail for seventy-five days; had lost his apartment, job, and truck and had been unable to pay child support.<sup>178</sup> Similarly, the ACLU's investigative report on debtors' prisons documents a homeless man arrested for stealing food worth \$39, fined \$339, and jailed, and a worker fined \$498 for possession of marijuana who was detained for five months for failure to pay a fine, among many other stories involving nonfelony offenses.<sup>179</sup> The State Court Processing Statistics Program omits these and thousands of other similar cases.<sup>180</sup> Moreover, there is considerable delay in reporting; the latest data available is currently from 2009.<sup>181</sup>

The federal government also sometimes prepares special reports that compile and analyze the county, state, and federal district court data that it collects. For example, a BJS report based on 1990–2004 State Court Processing Statistics describes overall pretrial release rates, the percentage of defendants detained at different amounts of money bail, and different pretrial release rates for black, white, and Hispanic defendants and male and female defendants for different types of crimes, among other statistics.<sup>182</sup> But this analysis does not compare these release rates among jurisdictions, thus failing to provide information about the counties, states, or federal district courts that might have the largest problems.

Some BJS data is state-specific. For example, through its Deaths in Custody Reporting Program, the BJS collects information from state prison and local jail death records and annual surveys, including, inter alia, the type

---

<sup>175</sup> See generally Alexandra Natapoff, *Misdemeanors*, 85 S. CAL. L. REV. 1313, 1313 (2012) (documenting the serious consequences of misdemeanors, which make up the “vast majority of U.S. convictions”).

<sup>176</sup> *Attorney General Eric Holder Speaks at the National Symposium on Pretrial Justice*, *supra* note 89.

<sup>177</sup> Laura Sullivan, *Bail Burden Keeps U.S. Jails Stuffed with Inmates*, NPR (Jan. 21, 2010), <http://www.npr.org/2010/01/21/122725771/Bail-Burden-Keeps-U-S-Jails-Stuffed-With-Inmates> [<https://perma.cc/5RBK-R8ZY>].

<sup>178</sup> *Id.*

<sup>179</sup> AM. CIVIL LIBERTIES UNION, *supra* note 36, at 6.

<sup>180</sup> See *supra* note 174–76 and accompanying text.

<sup>181</sup> *Data Collection: State Court Processing Statistics (SCPS)*, *supra* note 15 (noting that the latest data available is from 2009).

<sup>182</sup> See, e.g., COHEN & REAVES, *supra* note 27, at 1–6.



and cause of the death, offense type, and time served.<sup>183</sup> From this data it compiles reports that include, for example, prison and jail mortality rates by states; the percentages of deaths caused by particular conditions or actions, including suicide; and the average ages of inmates who died, among other data points.<sup>184</sup> However, this data does not provide county-level details that could help identify the jurisdictions with the most problematic practices, and, as with the State Court Processing Statistics Data, is often several years old by the time analysis of the information is completed and released.<sup>185</sup> The raw data behind the state-level statistics also is not readily available,<sup>186</sup> thus making it more difficult for other researchers to run their own analyses. As a first step toward improving the availability of criminal justice data—a project discussed in more detail in Part III—BJS should anonymize its raw data and make it electronically available to the public, eliminating the need for researchers to request the data on a case-by-case basis and provide relatively detailed justifications for accessing the data.

The Bureau also supports states' independent collection and statistical analysis of data through its Criminal Justice Data Improvement Program.<sup>187</sup> As part of this initiative BJS funds state Statistical Analysis Centers (SACs), which “collect, analyze, and report statistics on crime and justice.”<sup>188</sup> Recent SAC projects funded by BJS include, inter alia, an effort in Hawaii to automate certain information about supervised adult criminal defenders for easier compilation and statistical analysis<sup>189</sup>; an initiative in Minnesota to make “arrest, court, probation and correctional data” more publicly accessible<sup>190</sup>; and a study in certain New Mexico counties assessing the connection between pretrial detention and case processing times and the factors used in pretrial decisionmaking.<sup>191</sup> Thus, data collection is improving,

---

<sup>183</sup> MARGARET NOONAN ET AL., U.S. DEP'T OF JUSTICE, NCJ 248756, MORTALITY IN LOCAL JAILS AND STATE PRISONS, 2000–2013 - STATISTICAL TABLES 7–13 (Aug. 2015), <http://www.bjs.gov/content/pub/pdf/mljisp0013st.pdf> [<https://perma.cc/LJ2V-6SHE>].

<sup>184</sup> *Id.* at 3–4.

<sup>185</sup> *See id.* at 29–33. This time lag results in part from the BJS's careful use of statistical analysis and thorough approach to research.

<sup>186</sup> *See supra* notes 55–57 (describing how researchers must have their supervising institution's approval, such as an institutional review board's approval, in order to access raw criminal justice data from BJS).

<sup>187</sup> *Criminal Justice Data Improvement Program*, BUREAU JUST. STAT., <http://www.bjs.gov/index.cfm?ty=tp&tid=4> [<https://perma.cc/MX88-W2S6>].

<sup>188</sup> *State Justice Statistics Program*, BUREAU JUST. STAT., <http://www.bjs.gov/index.cfm?ty=tp&tid=48> [<https://perma.cc/H8PG-SZ7T>].

<sup>189</sup> *State Profiles*, BUREAU JUST. STAT., <http://www.bjs.gov/index.cfm?ty=tp&tid=481#Funding> [<https://perma.cc/W6B5-LD6V>].

<sup>190</sup> BUREAU OF JUSTICE STATISTICS, FY 2014 STATE JUSTICE STATISTICS (SJS) PROGRAM FOR STATISTICAL ANALYSIS CENTERS (SACS) SUMMARIES, [https://www.bjs.gov/content/pub/pdf/FY\\_2014\\_State\\_Justice\\_Statistics\\_Summaries.pdf](https://www.bjs.gov/content/pub/pdf/FY_2014_State_Justice_Statistics_Summaries.pdf) [<https://perma.cc/H8PG-SZ7T>].

<sup>191</sup> *Id.*

but SACs and the BJS fail to provide a consistent, easily accessible source of raw jurisdiction-by-jurisdiction criminal data such as the number of defendants released pretrial; which defendants received appointed counsel, and when; and the number of defendants jailed for a failure to pay court fines or similar fees.

For federal criminal justice data, the BJS also compiles a Compendium of Federal Justice Statistics that details the number of suspects arrested for a federal offense, describes the types and rates of offenses at arrest (for example, 46.3% of offenses in 2009 were immigration offenses)<sup>192</sup>, describes the percentage of arrests as allocated among judicial districts, outcomes of cases (plea, trial, or dismissal), sentence imposed, and mean prison terms, among other data.<sup>193</sup> There is also a large database of uniform, relatively thorough information on federal case information that can be accessed through the PACER system, which includes case summaries, docket entries, and documents from federal district, bankruptcy, and appellate court cases, including criminal cases.<sup>194</sup> Each night PACER collects “subsets of data” from all federal district courts through the PACER Case Locator, thus making the data available for searches.<sup>195</sup> Entities searching for information can search by the “nature of suit,” thus allowing for more broad-based searches than searches that are limited to party name or case number.<sup>196</sup>

Although PACER allows some bulk searches, it is even more expensive than many of the commercial services options. PACER users who search for criminal case data must pay 10 cents per page downloaded<sup>197</sup> in addition to the 10 cents per page charged for a PACER search, although “[t]he charge for any single document is capped at \$3, or 30 pages.”<sup>198</sup> Account holders who do not accrue charges of more than \$15.00 in a quarterly billing cycle do not owe PACER fees, although this fee exception is of limited use for bulk searches.<sup>199</sup>

---

<sup>192</sup> MARK MOTIVANS, U.S. DEP’T OF JUSTICE, NCJ 234184, FEDERAL JUSTICE STATISTICS, 2009, at 2 (Dec. 2011), <http://www.bjs.gov/content/pub/pdf/fjs09.pdf> [<https://perma.cc/X4LL-PC6Y>].

<sup>193</sup> *Id.* at 1–9.

<sup>194</sup> PACER, PACER USER MANUAL FOR ECF COURTS 24–28, <https://www.pacer.gov/documents/pacermanual.pdf> [<https://perma.cc/RPH4-PRX9>] (last updated Sept. 2014) (describing the information that PACER users can access).

<sup>195</sup> *Id.* at 14.

<sup>196</sup> *Id.* This search capability represents a change from previous practices. See brianwc, *What Is the “PACER Problem”?*, FREE L. PROJECT (Mar. 20, 2015), <https://free.law/2015/03/20/what-is-the-pacer-problem/> [<https://perma.cc/AG9T-68J6>] (“Unlike modern search engines that index documents and return responsive documents to queries, PACER users can only search for a litigant’s name or by using an already-known case docket number. This makes the system utterly unusable for general purpose research. You cannot search for a general word or phrase, like ‘asylum’ and hope to get back responsive documents. Instead one must already know the case one is looking for beforehand.”).

<sup>197</sup> PACER, *supra* note 194, at 15.

<sup>198</sup> *Id.* at 4.

<sup>199</sup> See PACER, ELECTRONIC PUBLIC ACCESS FEE SCHEDULE ¶8 (Dec. 2013), [https://www.pacer.gov/documents/epa\\_feesched.pdf](https://www.pacer.gov/documents/epa_feesched.pdf) [<https://perma.cc/Y33X-MUJQ>].

“[N]ew case reports” and “non-case-specific reports or searches”—the types of searches most relevant for those seeking bulk data—are not subject to the \$3 or thirty-page cap,<sup>200</sup> thus leading to potentially high charges for bulk searches. Further, although courts on a case-by-case basis may discretionarily waive PACER fees for nonprofit organizations, “individual researchers associated with educational institutions,” among other groups and individuals,<sup>201</sup> those requesting an exemption must show that the “exemption is necessary in order to avoid unreasonable burdens and to promote public access to information,” and individual researchers must show that the “defined research project” is “limited in scope” and “is not intended for redistribution on the internet.”<sup>202</sup> One recent Ninth Circuit order denying a nonprofit group’s request “for unlimited access to all court documents for twelve months” concluded that “[t]he amount of information sought is unprecedented in scope,” suggesting that large amounts of bulk data will be difficult to access.<sup>203</sup> The Ninth Circuit also has held that an order denying a PACER user’s request for a fee exemption is not a decision “of a judicial character” and is not appealable.<sup>204</sup>

The fact that PACER documents are free when viewed “at courthouse public access terminals”<sup>205</sup> is also of little use to individuals and groups seeking bulk records, as illustrated by Internet activist Aaron Swartz’s attempt to download the entire database: in response, the government temporarily halted the program and commenced an FBI investigation.<sup>206</sup> These and other limitations lead groups like the Free Law Project to conclude that “the average member of the public has no meaningful access to federal court records.”<sup>207</sup> Indeed, PACER and state and local court databases are primarily designed for attorneys and people looking for specific cases—not for individuals analyzing and coding bulk data to identify trends.

Finally, the Federal Sentencing Commission produces detailed monthly and quarterly reports on sentences submitted by all federal judges to the Commission, compiling cumulative data on within-recommended-range and out-of-range sentences, the primary offense and median sentence associated

---

<sup>200</sup> PACER, *supra* note 194, at 4.

<sup>201</sup> PACER, *supra* note 199, ¶ 9.

<sup>202</sup> *Id.*

<sup>203</sup> *In re* Application for Exemption from Elec. Pub. Access Fees by Public.Resource.org, No. 15-80056 (9th Cir. July 6, 2015) (order denying plaintiff’s motion for exemption).

<sup>204</sup> *In re* Application for Exemption from Elec. Pub. Access Fees by Gollan & Shifflett, 728 F.3d 1033, 1039 (9th Cir. 2013) (emphasis omitted) (quoting *United States v. Walton* (*In re Baker*), 693 F.2d 925, 927 (1982) (per curiam)).

<sup>205</sup> PACER, *supra* note 199, ¶ 8.

<sup>206</sup> John Schwartz, *An Effort to Upgrade a Court Archive System to Free and Easy*, N.Y. TIMES (Feb. 12, 2009), [http://www.nytimes.com/2009/02/13/us/13records.html?\\_r=2](http://www.nytimes.com/2009/02/13/us/13records.html?_r=2) [<https://perma.cc/G7QB-YHYS>].

<sup>207</sup> brianwc, *supra* note 196.

with that offense, and other statistics.<sup>208</sup> The Commission has also responded to demands for new types of information, increasing “[t]he types of data reported” based on specific data requests.<sup>209</sup> Much of this data is summary data, however, and lacks the raw data needed for bulk analysis.

### 3. *Commercial Services*

As described above, government collection and reporting of bulk criminal justice data is incomplete at best. However, one could imagine a service in which a private company sent individual employees to hundreds of courthouses, police departments, corrections departments, and clerks offices to collect and record extensive data from publicly accessible files; compiled this information, and made this information accessible (for a fee) in a uniform, searchable format that allowed for the type of extensive statistical analysis typical of the big data movement. Indeed, there are numerous commercial services that do this very task, such as LexisNexis Risk Solutions<sup>210</sup> and US Investigations Services, LLC (formerly a government agency, now privatized),<sup>211</sup> among many others.<sup>212</sup> Commercial criminal justice data companies, like any individual searching for information, still must work with

---

<sup>208</sup> See generally U.S. SENTENCING COMM’N, PRELIMINARY QUARTERLY DATA REPORT: 4TH QUARTER RELEASE (2015), [http://www.ussc.gov/sites/default/files/pdf/research-and-publications/federal-sentencing-statistics/quarterly-sentencing-updates/USSC-2015\\_4th\\_Quarterly\\_Report.pdf](http://www.ussc.gov/sites/default/files/pdf/research-and-publications/federal-sentencing-statistics/quarterly-sentencing-updates/USSC-2015_4th_Quarterly_Report.pdf) [<https://perma.cc/NXW6-55SH>].

<sup>209</sup> *Id.* intro.

<sup>210</sup> See *LexisNexis Risk Solutions*, LEXISNEXIS, <http://www.lexisnexis.com/risk/?gclid=CLvwo52348oCFQFkhgodGrwMIA> [<https://perma.cc/6PMV-3EFQ>]. This service was previously ChoicePoint, Inc. See *Acquisition of ChoicePoint Inc. Completed*, RELX GROUP (2008), <http://www.relx.com/mediacentre/pressreleases/2008/Pages/AcquisitionofChoicePointIncCompleted.aspx> [<https://perma.cc/A2YG-W9H2>]; *Reed Elsevier to Acquire ChoicePoint for \$3.6 Billion*, N.Y. TIMES (Feb. 21, 2008), <http://www.nytimes.com/2008/02/21/technology/21iht-reed.4.10279549.html> [<https://perma.cc/F9VS-9HHZ>] (“Reed, based in London, said that combining ChoicePoint with its LexisNexis information service and its Analytics group would create a risk-management business with \$1.5 billion in revenue and a leading position in a fast-growing market.”). ChoicePoint, Inc. allowed breaches of its data and failed to adequately screen the individuals accessing data, thus allowing identity theft. See *ChoicePoint Settles Data Security Breach Charges; to Pay \$10 Million in Civil Penalties, \$5 Million for Consumer Redress*, FED. TRADE COMMISSION (Jan. 26, 2006), <https://www.ftc.gov/news-events/press-releases/2006/01/choicepoint-settles-data-security-breach-charges-pay-10-million> [<https://perma.cc/WN54-W4LU>]. In 2006 it reached a settlement with the Federal Trade Commission for \$10 million in civil penalties and \$5 million in consumer redress. *Id.* LexisNexis Risk Solutions also acquired what was previously Rapsheets.com. See *Rapsheets*, KNOWX, <https://www.knowx.com/rapsheets/> [<https://perma.cc/F3N8-FG38>].

<sup>211</sup> See *Company Overview of US Investigations Services, LLC*, BLOOMBERG (Jan. 4, 2017), <http://www.bloomberg.com/research/stocks/private/snapshot.asp?privcapId=115635> [<https://perma.cc/43NZ-KL7F>].

<sup>212</sup> SEARCH, *supra* note 143, at 7–8.

government officials to obtain this information.<sup>213</sup> But because these companies' mission is to collect data, they are more likely to be familiar with the procedures for collecting the information, and thus more efficient at the collection effort.<sup>214</sup> Further, they have the resources to purchase bulk data in the courts and other institutions that allow these types of purchases.<sup>215</sup>

Despite the potential for commercial services to collect and provide—for a price—data from across jurisdictions, the types of information are limited by the market for them, or lack thereof. Some of the most common information collected and provided is for the purposes of background searches by employment agencies and other organizations.<sup>216</sup> This data includes, for example, information on “the arrest (or notice to appear in lieu of arrest); detention; indictment or other formal criminal charge (and any conviction, acquittal, or other disposition arising therefrom); sentencing; correctional supervision; or release.”<sup>217</sup> Some of this data could assist criminal justice reform efforts by, for example, indicating whether the charge related to a failure to pay a probation fee and whether the individual was sentenced to imprisonment for this failure to pay. But because the data tends to be focused on understanding an individual's criminal justice record, it lacks information such as the time between arrest and appointment of counsel and the length of pretrial detention.

In addition to not collecting many of the types of criminal justice information most valuable to scholars, policymakers, and reform groups, commercial services that collect and digitize thousands of court and police records also charge fees that make this information inaccessible to these groups absent a generous grant or other funding.<sup>218</sup> For example, KnowX—part of Lexis's service—charges \$9.95 per background check or \$25.00 for an “[a]dvanced” check.<sup>219</sup> Searchers can also obtain a day-pass for \$24.95<sup>220</sup> or a thirty-day pass with unlimited background searches for \$99.95.<sup>221</sup> There are millions of criminal defendants in the United States,<sup>222</sup> and even if the background check data most frequently collected by commercial services had

---

<sup>213</sup> *Id.* at 2 n.5.

<sup>214</sup> *Id.*

<sup>215</sup> *Id.*

<sup>216</sup> *Id.* at 19.

<sup>217</sup> *Id.* at 5.

<sup>218</sup> *See, e.g., Standard: Pricing, KNOWX*, <https://www.knowx.com/statmnts/priceinfo.jsp> [<https://perma.cc/2QLL-4LSQ>].

<sup>219</sup> *Id.*

<sup>220</sup> *Id.*

<sup>221</sup> *Background Check Unlimited Search Subscription*, KNOWX, <https://www.knowx.com/subscriptions/k-31pop-up.jsp> [<https://perma.cc/SN7Z-4Z6W>].

<sup>222</sup> *See* ADMIN. OFFICE OF THE U.S. COURTS, TABLE D-1: CRIMINAL DEFENDANTS COMMENCED, TERMINATED, AND PENDING (INCLUDING TRANSFERS), DURING THE 12-MONTH PERIOD ENDING JUNE 30, 2015 (June 2015), [http://www.uscourts.gov/sites/default/files/d01djun15\\_0.pdf](http://www.uscourts.gov/sites/default/files/d01djun15_0.pdf) [<https://perma.cc/W8Q2-UNKR>] (showing that 78,592 criminal defendants commenced during this time period).

the data points most critical to revealing problems in priority areas, data collection through this method could be cost prohibitive. Conceivably, however, commercial data collection services could be induced to provide anonymized (and thus valueless to their principal customers) bulk data to researchers at a reduced cost. In the absence of a more comprehensive solution, this is an avenue worth exploring.

Considered collectively, the criminal justice data available is surprisingly limited or expensive to access. A few states, like New York<sup>223</sup> and Pennsylvania,<sup>224</sup> have begun to collect more detailed data, to ensure that the data is consistent across municipalities and counties in the state, and to make this data more accessible. Further, in the many cases where courts are not providing adequate information, some enterprising companies are digging up this information themselves, traveling to courtrooms around the country and digitizing records.<sup>225</sup> One might expect that a variety of incentives would push courts and companies to collect more of this information and at least provide it for a fee, using the money to fund much-needed court services and staff or to simply produce profits for a company. But the nature of the information prevents the formation of strong markets in this area, as discussed below.<sup>226</sup>

### B. *Inadequate Incentives for Data Production*

Given the huge volume of data now electronically available, it is perhaps surprising that we still do not know how many courthouses in how many counties and states deny the public access to the courtroom,<sup>227</sup> fail to ask a defendant whether he or she is indigent and therefore needs counsel,<sup>228</sup> apply bail schedules that cause the automatic detention of numerous defendants,<sup>229</sup> or jail defendants who are unable to pay parole fees or child support.<sup>230</sup> One might expect, that even if this information is not currently systemically collected and reported, someone would be likely to step in to fill this gap. But

---

<sup>223</sup> See *supra* notes 151–53 and accompanying text.

<sup>224</sup> See *supra* note 166 and accompanying text.

<sup>225</sup> See *supra* notes 210–11 and accompanying text.

<sup>226</sup> See *infra* Part III.B.

<sup>227</sup> The limited data available in this area tends to be from a limited number of lawsuits filed against individual counties. See *infra* notes 246–49 and accompanying text.

<sup>228</sup> Currently, there is only very broad data on the failings of indigent representation and on a limited number of individual jurisdictions. See *supra* notes 103–14 and accompanying text.

<sup>229</sup> Once again, there is broad data on this practice, suggesting that 64% of state and county courts use bail schedules, and data from a limited number of jurisdictions. See PRETRIAL JUSTICE INST., *supra* note 88, at 7.

<sup>230</sup> As with the other criminal justice problems discussed in this Article, there are only broad surveys with anecdotes of the likely prevalence of debtors' prisons and limited lawsuits in certain jurisdictions. See *supra* notes 36, 119–22 and accompanying text.

this information—while it could be, and sometimes is<sup>231</sup>—transformed into a marketable good—lacks a significant market. And this barrier is compounded by strong resistance from many county and state courthouses to the prospect of digitizing and widely disseminating information that could reveal negative characteristics of their criminal justice systems.

### 1. *A Public Good Without a Market Solution*

Data about the millions of defendants who annually encounter some part of the local, state, or federal criminal justice system on an annual basis, which is a public good, encounters market failures that have long been associated with information production and dissemination more generally.<sup>232</sup> As many copyright owners have discovered in the digital age, if information is produced and disseminated, it is difficult to prevent others from accessing and benefitting from this data.<sup>233</sup> Thus, electronically accessible criminal justice data is a classic nonrival public good for which no user has incentive to pay, and which suffers from chronic underproduction.<sup>234</sup> In many contexts, this problem has been overcome by rapidly declining costs of information production and dissemination,<sup>235</sup> but many courts still view data production as costly.

Even when information gathering remains costly, firms can maintain demand by restricting access. Westlaw and LexisNexis, for example, are able to charge significant fees for access to their databases (although judicial opinions are increasingly becoming available for free online)<sup>236</sup>, because bulk downloads are not permitted. However, in order for firms to be incentivized to

---

<sup>231</sup> See *supra* Part III.A.3 (describing commercial services that send staff to courthouses around the country to collect and digitize certain data relating to criminal cases).

<sup>232</sup> See James Gibson, *Re-Reifying Data*, 80 NOTRE DAME L. REV. 163, 174 (2004) (“The developer of a database may expend substantial time and effort in researching and assembling the data. Once a single copy of the database is published, however, the developer has little ability to exclude anyone from making and distributing competing copies, and the price it can demand will therefore quickly be driven down toward the marginal cost of producing copies.”).

<sup>233</sup> See *id.*

<sup>234</sup> See *id.* at 172–73 (observing that “information is a public good” and describing its nonrival and nonexcludable characteristics); cf. David Thacher, *The Rise of Criminal Background Screening in Rental Housing*, 33 LAW & SOC. INQUIRY 5, 22 n.12 (2008) (“[C]riminal history information . . . is a collective good because (like information goods in general) it is essentially nonrival: once it has been assembled for one user, the marginal cost of providing it to additional users is very small.”).

<sup>235</sup> See, e.g., Kristin Madison, *Regulating Health Care Quality in an Information Age*, 40 U.C. DAVIS L. REV. 1577, 1579 (2007) (noting a decline of “information failure” in healthcare due to “[r]ecent innovations in the production, analysis, and dissemination of information”).

<sup>236</sup> See, e.g., Legal Info. Inst., *Federal Law: Judicial Opinions*, CORNELL U. L. SCH., <https://www.law.cornell.edu/federal/opinions.html> [<https://perma.cc/TG2K-KZLA>].

collect and digitize the information, and put up access barriers, there must be a demand for the data itself. For instance, the number of background checks, conducted by public and private employers has grown substantially in recent years,<sup>237</sup> creating a profitable market for firms like KnowX.<sup>238</sup> But the type of information needed for criminal justice reform, such as who is detained pretrial and why, how many indigent defendants are denied counsel and at what stage, and the number and types of defendants jailed due to a failure to pay criminal fees, lacks a similar market. Rather, this data is demanded by a small, select group of scholars and reformers, many of whom lack the resources to pay hefty fees for the thousands of data points needed to produce good empirical information in this area.<sup>239</sup> Indeed, their work itself is a species of often-underproduced public good. The lack of commercial demand for the information, and relatively low ability to pay on the part of those who could benefit from the information, causes the market not to provide it.

As seen above, governments also largely fail to provide access to these sorts of data as well,<sup>240</sup> albeit for different reasons. Inertia, no doubt, plays a part. So, too, do concerns about defendants' privacy. Indeed, some state policies expressly prohibit courts' dissemination of electronic data, or disallow bulk data searches of court data, out of concern that individuals and commercial services will use the information for illicit purposes, although, as described below, these concerns can be addressed by stripping bulk data of identifying information.<sup>241</sup> Less laudably, it may be that some courts resist providing broader access to their information because of the rent payments generated by charging for it. For example, federal courts have used the fees from PACER for projects such as providing flatscreen monitors for each individual juror in certain courtrooms and embedding speakers within courtroom benches and panels to enhance courtroom audio.<sup>242</sup> Further, in some cases, courts must do more than simply make their electronic courthouse data available more broadly (with privacy protections). As with many information systems, courts have long followed particular practices when it

---

<sup>237</sup> See Daniel J. Solove & Chris Jay Hoofnagle, *A Model Regime of Privacy Protection*, 2006 U. ILL. L. REV. 357, 375 (noting the rise of background checks even for "menial" jobs).

<sup>238</sup> See *supra* notes 220–21 and accompanying text.

<sup>239</sup> See, e.g., *supra* note 76 and accompanying text (describing an extensive academic data project that required access to state court cases requiring a federal grant); *supra* notes 59–66 and accompanying text (describing an effort to collect pretrial data funded by a foundation—a project that was only made possible through the generous donations of wealthy individuals); *supra* notes 18, 118–22 (describing cases filed on a jurisdiction-by-jurisdiction basis and studies conducted through interviews and staff sifting through court files in order to begin to identify jurisdiction specific problems).

<sup>240</sup> See *supra* notes 145–66, 181–205 and accompanying text.

<sup>241</sup> See *supra* note 143 and accompanying text.

<sup>242</sup> Schultze, *supra* note 45.



comes to collecting information,<sup>243</sup> and asking them to collect new types of information and make it available in an anonymized, uniform manner creates costs and coordination problems. Thus, lingering costs to information production—whether real or perceived—contribute to an ongoing information failure in the criminal justice context. But state and local courts’ resistance to calls to produce more and more uniform data may also stem from a desire to avoid unwanted scrutiny or oversight.

## 2. Local Opposition to Data Collection and Disclosure

Some state and local governments may oppose any effort that supports detailed analysis—research that could expose pathologies within certain jurisdictions’ criminal justice systems, such as systemic denial of counsel to indigent defendants. While some governments may welcome the opportunity to identify and correct problems, others may be wary of increased exposure to criticism and litigation. For governments violating well-established rights, exposure can lead to rapid change. For example, in a case initiated by Equal Justice Under Law attorneys, a U.S. district court in Alabama noted that the City of Clanton, Alabama “utilized a generic bail schedule for new misdemeanor arrests, a schedule from which it did not deviate: \$500 per charge (and \$1,000 per DUI charge),” and required “payment upfront,” thus causing those unable to pay to be held in jail until the next court date.<sup>244</sup> However, City officials revised their bail policies before the court issued its opinion.<sup>245</sup> The officials also hastily implemented a policy allowing public access to the city’s municipal courtroom<sup>246</sup> in response to allegations that it regularly denied access.<sup>247</sup> Similarly, in Williamson County, Texas, after plaintiffs alleged that the County denied open access to its courtrooms and regularly failed to provide counsel to indigent defendants,<sup>248</sup> the County purported to change its policies in an unsuccessful attempt to have the case dismissed as moot.<sup>249</sup>

Many state and county courts are also likely to view federal data collection as yet another intrusion into their criminal justice systems and a threat to comity.<sup>250</sup> Indeed, this might explain, in part, why only relatively few states

---

<sup>243</sup> See *supra* note 142 (describing certain state courts’ rules for record retention and public access).

<sup>244</sup> *Jones v. City of Clanton*, No. 2:15cv34–MHT, 2015 WL 5387219, at \*1 (M.D. Ala. Sept. 14, 2015).

<sup>245</sup> *Id.* at \*3.

<sup>246</sup> *Id.* at \*2.

<sup>247</sup> *Id.* at \*1.

<sup>248</sup> *Williamson County v. Heckman*, 368 S.W.3d 1, 15 (Tex. App. 2010), *rev’d in part*, 369 S.W.3d 137 (Tex. 2012).

<sup>249</sup> See *id.* at 15–16.

<sup>250</sup> Cf. Samuel R. Wiseman, *What Is Federal Habeas Worth?*, 67 FLA. L. REV. 1157, 1188 (2015) (describing scholars’ arguments for federal deference to state habeas decisions to protect the dignity and respect the interests of states).

have signed on to some of the Federal Bureau of Investigation's numerous initiatives encouraging uniform data entry for crime incidents, crime histories, and other data.<sup>251</sup> Although more than 6,000 local, state, and tribal agencies participate in the National Incident-Based Reporting System (NIBRS)—a law enforcement system with detailed data about crimes and the characteristics of individuals committing the crime<sup>252</sup>—only fifteen states directly input crime incidents into the system, and an additional eighteen states have systems that are certified by NIBRS and compatible with it.<sup>253</sup> As shown by congressional efforts to force more uniform data collection and reporting, such as information on individuals not licensed to carry firearms, the federal government often must cajole state and local governments to cooperate by expanding the grant money offered to them.<sup>254</sup>

The criminal justice system is a black, or at least gray, box because of a variety of disincentives for its production and dissemination, ranging from information failures<sup>255</sup> to local and state resistance to efforts toward uniform collection and electronic dissemination of data.<sup>256</sup> Certain local governments and states, however, have begun to generate and make available the types of information that are needed to pinpoint the parts of the criminal justice system most in need of reform, and to identify the jurisdictions with the largest problems.<sup>257</sup> The following Part suggests ways of improving information production and access.

#### IV. CLOSING THE DATA GAP

Criminal justice data is far behind other fields in terms of the availability of uniform, electronic data for sophisticated analysis. As Professor Daniel Esty noted in the environmental context in 2004—when electronic monitoring and collection of pollution data was already rapidly expanding—areas with particularly broad information gaps often require a centralized solution for the purposes of enhancing efficiency, improving economies of scale, and mediating inconsistencies among potential data gathering systems.<sup>258</sup> This centralized entity need not always collect information—indeed, sometimes smaller-scale collection, provided that it is coordinated by a centralized entity,

---

<sup>251</sup> See *infra* Part IV.

<sup>252</sup> UNIF. CRIME REPORTING (UCR) PROGRAM, NATIONAL INCIDENT-BASED REPORTING SYSTEM (NIBRS), NIBRS PARTICIPATION BY STATE (2013), <https://ucr.fbi.gov/nibrs/2013/resources/nibrs-participation-by-state> [<https://perma.cc/65X8-GVUK>].

<sup>253</sup> *Id.*

<sup>254</sup> See *infra* note 287 and accompanying text.

<sup>255</sup> See Gibson, *supra* note 232, at 172–74.

<sup>256</sup> See generally Jones v. City of Clanton, No. 2:15cv34–MHT, 2015 WL 5387219 (M.D. Ala. Sept. 14, 2015).

<sup>257</sup> See *State Profiles*, *supra* note 189.

<sup>258</sup> Esty, *supra* note 8, at 143–44.

is the least expensive option.<sup>259</sup> But the federal government is best situated to create a uniform system for data collection and reporting, regardless of whether private entities, states, local governments, or federal agencies collect the information.

#### A. Improved Data Collection System

As described above, some states have already unified their information collection systems for criminal defendants and their interactions with the criminal justice system, ensuring that all municipal and county courts provide basic electronic data about defendant demographics, criminal history, arrest, types and numbers of charges, and case dispositions.<sup>260</sup> These states also provide information on how to uniformly input data and correct incorrect data, and they conduct analyses to determine whether data is in fact being inputted correctly across jurisdictions.<sup>261</sup> Additionally, some counties and states provide detailed docket information and case summaries that contain certain data on defendant demographics, charges, appointment of counsel, case disposition, and sentences, but the data collected and made available varies widely.<sup>262</sup> Individual courts are similarly varied, with the most closed courts providing no electronically accessible information and thus requiring on-site data collection.<sup>263</sup> The federal government, as described above, collects detailed information about felony defendants from forty of the largest seventy-five counties, although this data is from limited time periods within each data year, it does not include certain important pretrial information, and researchers need permission to access it.<sup>264</sup> The federal government also has conducted periodic surveys of obstacles to effective indigent defense, describing the counties with the highest caseloads, for example, and generalized data about a lack of adequate access to investigative resources.<sup>265</sup> But this data, too, lacks specifics on the percentages of indigent defendants denied counsel and why. In sum, despite the enormous advances in information technology made over the last few decades, much criminal justice data remains difficult to access, if it exists at all. Modest changes could produce significant benefits. BJS and New York, for example, could make anonymized versions of their databases available to the public, and PACER's fee structure and search features could be improved.<sup>266</sup> And as noted above, providers of background checks might be willing to allow researchers access to a version of their databases with

---

<sup>259</sup> *Id.* at 145.

<sup>260</sup> *See supra* notes 153–59 and accompanying text.

<sup>261</sup> *See supra* notes 161–64 and accompanying text.

<sup>262</sup> *See supra* note 150 and accompanying text.

<sup>263</sup> *See supra* note 150 and accompanying text.

<sup>264</sup> *See supra* notes 54–57 and accompanying text.

<sup>265</sup> *See supra* notes 104–13 and accompanying text.

<sup>266</sup> *See supra* Part III.A.2.

identifying information removed.<sup>267</sup> But a more comprehensive solution would have even greater potential to aid research and reform efforts.

This Part accordingly proposes that the federal government provide an electronic platform through which counties and states would input data for each defendant, including: 1) demographic data such as gender, race, age, and individual characteristics identified during a bail hearing, such as income and community ties; 2) arrest and arraignment dates and date of the charge; 3) date on which counsel was appointed, and reasons for a failure to appoint counsel if no counsel was appointed; 4) date of the bail hearing; 5) pretrial detention, bail, or pretrial release imposed, including conditions associated with release; 6) whether or not the defendant appeared for trial or committed pretrial crimes; 7) date of case disposition and type of adjudication, including a plea, conviction, diversion, or determination of not guilty; 8) the sentence imposed; 9) and parole conditions imposed. For states like New York and Pennsylvania that already have unified electronic databases,<sup>268</sup> the federal government could set uniform standards for data coding and anonymizing, allowing the incorporation of those databases without duplication of effort. This data would allow reformers, judges, police, policymakers, and researchers to identify how long defendants sit in jail or are subjected to electronic monitoring or other conditions while awaiting trial; the jurisdictions that appoint counsel too late in a case (or never), and how often this happens; and the types of charges that lead to imprisonment, including charges associated with a failure to pay court fines and other monetary obligations. It may be that other, similar types of information would be useful; this Article targets the basic data necessary to inform the three criminal justice efforts introduced in Part I, but additional data points are possible.

County, state, and federal courts—which already collect much of this data, albeit not uniformly or consistently—are the logical entities to input this data and provide it to a centralized database, similar to the unified approach that New York uses for its county and municipal courts.<sup>269</sup> This database should have uniform data input fields containing the information identified above, and individual courts' data entries should be directly linked electronically to this database, thus allowing automated transfer of data inputs from individual courts to the centralized database. Also drawing from the New York approach, there should be uniform rules for inputting data—such as inputting dates in the same format—and for correcting data.<sup>270</sup> A computer program should also scan the data and flag anomalies, such as the data of case disposition being earlier than the date of arraignment. To protect defendants' privacy, all names, addresses, social security and driver's license numbers, telephone numbers, and other personal identifying characteristics should be omitted from the

---

<sup>267</sup> See *supra* Part III.A.3.

<sup>268</sup> See *supra* notes 153–66 and accompanying text.

<sup>269</sup> See *supra* notes 153–64 and accompanying text.

<sup>270</sup> See *supra* notes 161–64 and accompanying text.

database—similar to the system already followed in states like Idaho.<sup>271</sup> Further, the public should have free, open access to this database without having to request access and provide documentation of institutional approval for the research.<sup>272</sup>

A comprehensive database is a lofty goal, but, once its architecture is in place, it could be built incrementally as states and local governments choose to opt in; even a partial database would be of considerable value. Similarly, even without federal participation or standardization, states can, with varying degrees of effort, make their own data significantly more accessible, a worthwhile endeavor in itself.

### B. *Federal Involvement*

The process of persuading county and state courts to record the same types of data, and in the same format, will be a complex task. The logical entity for overcoming the high coordination and resource-based hurdles to this effort is the federal government—and particularly the Bureau of Justice Statistics—which already collects some of this data through surveys.<sup>273</sup>

Federal leadership in creating and maintaining the criminal justice database proposed here will be necessary for reasons of political economy and the collective-action challenges identified in Part II.B. The federal government could offer incentives to persuade subfederal entities to agree to these perceived intrusions, such as providing enhanced funding for police forces or other criminal justice efforts for those jurisdictions that agree to provide data in uniform ways. Indeed, the government in the past has offered incentive funding to states that provide information about individuals ineligible to purchase guns and that improve and share their criminal justice data.<sup>274</sup> Further, because the BJS already collects local and state data through surveys, the use of an automated information collection system might reduce burdens on state and local governments, eliminating the need to systematically collect and report data for these surveys.

Federal intervention in collecting data is also important to create uniformity. A wholly bottom-up effort, in which thousands of court systems developed their own collection and reporting efforts would—while still an improvement over the status quo—make cross-jurisdiction analysis difficult, if not impossible. Although a federal lead is necessary to overcome these problems, state and local input will also be important to incentivize more subfederal jurisdictions to participate in the database. The BJS or a similar federal agency that formulates the database and works with state and local

---

<sup>271</sup> See *Idaho Supreme Court Data Repository*, *supra* note 46.

<sup>272</sup> This would make the database more accessible than current repositories, such as the Bureau of Justice Statistics' data from counties and states. See *supra* note 54 and accompanying text.

<sup>273</sup> See *supra* notes 173–75, 183–84, 192–93 and accompanying text.

<sup>274</sup> See *infra* notes 287–300.

governments to obtain information for the database should examine current subfederal data collection practices identified in Part I of this Article and conform the database as closely as possible to these existing practices. The more that a federal database matched existing subfederal practices, the more likely that states, counties, and other local governments would be to provide information; this would also substantially reduce burdens on those jurisdictions that chose to participate in the database.

The federal government has a long history of coordinating state criminal justice information collection efforts,<sup>275</sup> and its expertise and superior resources in this area could be leveraged to create the criminal justice database proposed here. The government has historically formed federal databases that collected state-based criminal information (particularly arrest and other criminal history information),<sup>276</sup> examined existing state data collection efforts to determine how national databases could best harness information from states,<sup>277</sup> and required uniform state reporting and linkages of state data with federal databases.<sup>278</sup>

In one example of an extensive federal data collection effort in the criminal context, the Brady Act of 1993 directed the Attorney General to create a National Instant Criminal Background Check System (NICS),<sup>279</sup> which firearms importers, dealers, and licensees were required to search before transferring a firearm to an unlicensed person.<sup>280</sup> Congress directed the federal government to investigate existing state criminal records systems to determine how quickly states would likely be able to provide data to a national on-line system.<sup>281</sup> Further, Congress required the Attorney General to develop the necessary computer system to link state criminal history background checks with the federal NICS, thus allowing for automated transfer of data between the two systems.<sup>282</sup>

As originally formed, NICS did not provide the FBI with “automated access to complete information from the States” regarding individual possession of firearms, and “millions of criminal records” lacked “critical data, such as arrest dispositions, due to data backlogs.”<sup>283</sup> Congress identified the primary obstacles to this data effort as arising from the states, including

---

<sup>275</sup> See BUREAU OF JUSTICE STATISTICS, *supra* note 190, at ix.

<sup>276</sup> See *supra* notes 173–75, 188 and accompanying text.

<sup>277</sup> BUREAU OF JUSTICE ASSISTANCE & NAT’L INST. OF JUSTICE, STANDARD FUNCTIONAL SPECIFICATIONS FOR LAW ENFORCEMENT RECORDS MANAGEMENT SYSTEMS 1 (2003), [http://it.ojp.gov/documents/LEITSC\\_Law\\_Enforcement\\_RMS\\_Systems.pdf](http://it.ojp.gov/documents/LEITSC_Law_Enforcement_RMS_Systems.pdf) [<https://perma.cc/L2ES-BCC7>].

<sup>278</sup> See *infra* notes 279–99 and accompanying text.

<sup>279</sup> Brady Handgun Violence Prevention Act, Pub. L. No. 103-159, § 103(b), 107 Stat. 1536 (1993) (codified as amended at 18 U.S.C. §§ 921–923, 925(a) (2012)).

<sup>280</sup> *Id.* § 102 (b)(t)(1).

<sup>281</sup> *Id.* § 103 (a)(2).

<sup>282</sup> *Id.* § 103(c)(2).

<sup>283</sup> NICS Improvement Amendments Act of 2007, Pub. L. No. 110-180, § 2(2)–(4), 121 Stat. 2259 (codified as amended at 18 U.S.C. § 922 (2012)).

inadequate updates and availability of state records and a lack of automated access to many of these records.<sup>284</sup> It accordingly recommended that states make the data federally available in a more “usable” format and computerize and automate their records with the assistance of the federal government.<sup>285</sup> Congress incentivized states to make their records on firearm possession eligibility electronically accessible,<sup>286</sup> promising that states would not have to provide any matching funding of federal money already provided for these types of efforts.<sup>287</sup> To receive this incentive, states also were required to update their records when certain criteria on firearm possession ineligibility expired, thus ensuring that state data was up-to-date and uniform.<sup>288</sup>

Another example of federal involvement in state information collection and electronic reporting is the FBI’s Interstate Identification Index (III), which contains criminal information on individuals, and in which all states participate.<sup>289</sup> The index includes identification information such as “name, birth date, race and gender” for all “persons arrested for fingerprintable [state or federal] felonies and misdemeanors.”<sup>290</sup> Further, the FBI maintains a centralized database of state and federal data designed to help “criminal justice professionals apprehend fugitives, locate missing persons, recover stolen property, and identify terrorists.”<sup>291</sup> This National Crime Information Center database includes numerous electronic records, from those in the National Sex Offender Registry to missing persons, suspected terrorists, and identity theft cases.<sup>292</sup>

At the law enforcement level, many states also have agreed to a uniform data reporting system organized by the federal government.<sup>293</sup> Through the NIBRS, more than 6,000 local, state, tribal, and federal agencies report crime data into the system,<sup>294</sup> including data such as “assaults on law enforcement officers, offenses in which weapons were involved, drug/narcotic offenses, hate crimes, domestic and familial abuse including elder abuse, juvenile crime, gang-related crime, parental abduction, organized crime, and pornography, as well as arrest data related to driving under the influence.”<sup>295</sup> Agencies also

---

<sup>284</sup> *Id.* § 2(5).

<sup>285</sup> *Id.* § 2(6).

<sup>286</sup> *Id.* § 102(c).

<sup>287</sup> *Id.* § 102(a).

<sup>288</sup> *Id.* § 102(c).

<sup>289</sup> *Compiling & Disseminating Criminal History Record Information*, SEARCH, <http://www.search.org/solutions/criminal-history-records/compiling-disseminating-criminal-history-record-information/> [https://perma.cc/6M4X-X3J5].

<sup>290</sup> *Id.*

<sup>291</sup> *National Crime Information Center (NCIC)*, FBI, <https://www.fbi.gov/services/cjis/ncic> [https://perma.cc/MW8W-982X].

<sup>292</sup> *Id.*

<sup>293</sup> UNIF. CRIME REPORTING (UCR) PROGRAM, *supra* note 252.

<sup>294</sup> *Id.*

<sup>295</sup> *2013 National Incident-Based Reporting System*, FBI: UCR, <https://www.fbi.gov/about-us/cjis/ucr/nibrs/2013> [https://perma.cc/QC7F-84L6].

enter offender-specific information such as drug and alcohol use “before the incident and whether the offender used computer equipment to perpetrate the crime.”<sup>296</sup> States can directly input information into the national database or into a state program that is certified as meeting the federal standards.<sup>297</sup> The federal government ensures uniform data input by carefully defining twenty-three offense categories containing forty-nine “specific crimes,” and all individualized systems that participate in NIBRS are automated.<sup>298</sup>

These and other efforts to collect and synthesize state criminal history information have involved extensive work by the federal government to make state reporting more consistent, most of which is funded, directed, or administered by the BJS, the Bureau of Justice Assistance, and an information-based division of the FBI. For example, these bureaus have supported uniform development of local and state records management systems maintained by law enforcement, recommending standard practices such as automatic submittal of data to external sources, use of codes whenever possible, and validation of data entry for each data field included.<sup>299</sup> The BJS also provides annual grants to states and local governments to “improve the quality, timeliness, and immediate accessibility of criminal history records and related information” and to integrate and link information from all components of the criminal justice system, such as prosecutorial, police, court, and jail and prison records.<sup>300</sup> BJS also worked with the FBI’s Criminal Justice Information Services Division and a nonprofit organization of states to develop voluntary performance standards for improving state criminal history repositories and to make them more uniform.<sup>301</sup>

The costs to state and local governments associated with updating their information collection and reporting efforts and unifying these efforts could also be supported by federal funds. A good deal of federal criminal information-related funding currently supports improved and more uniform data for criminal background checks.<sup>302</sup> Although this data is important for crime prevention, gun control, and other efforts, the government has recently expanded its attention to systemic criminal justice problems, particularly

---

<sup>296</sup> *Id.*

<sup>297</sup> UNIF. CRIME REPORTING (UCR) PROGRAM, NATIONAL INCIDENT-BASED REPORTING SYSTEM (NIBRS), *METHODOLOGY* (2013), <https://fbi.gov/about-us/cjis/ucr/nibrs/2013/resources/methodology> [<https://perma.cc/EN6U-D974>].

<sup>298</sup> *Id.*

<sup>299</sup> BUREAU OF JUSTICE ASSISTANCE & NAT’L INST. OF JUSTICE, *supra* note 277, at 1.

<sup>300</sup> *National Criminal History Improvement Program*, BUREAU JUST. STAT., <http://www.bjs.gov/index.cfm?ty=tp&tid=47> [<https://perma.cc/Y2LY-37R3>].

<sup>301</sup> *State Repository Records & Reporting Quality Assurance Program (QAP)*, SEARCH, <http://www.search.org/solutions/criminal-history-records/state-repository-record-s-and-reporting-quality-assurance-program-qap/> [<https://perma.cc/6MPM-8P6D>].

<sup>302</sup> *See, e.g., State Profiles*, BUREAU JUST. STAT., <http://www.bjs.gov/index.cfm?ty=tp&tid=491#promising> [<https://perma.cc/PL5G-U7GM>] (showing awards for improving state record reporting for the National Instant Criminal Background Check System as increasing from a total of \$2,506,731 in 2009 to \$22,695,054 in 2015).



through its funding of state SACs, which “collect, analyze, and report statistics on crime and justice.”<sup>303</sup> Not all of this funding, which is provided to the SACs as part of the Criminal Justice Data Improvement Program,<sup>304</sup> supports the type of data needed for criminal justice reform,<sup>305</sup> but some does. For example, as introduced above, two recent grants support an effort in Minnesota to make “arrest, court, probation and correctional data” more publicly accessible and useable; and a study in certain New Mexico counties assessing the connection between pretrial detention and case processing times and the factors used in pretrial decisionmaking.<sup>306</sup> Some additional shifting of funds to the criminal justice database might be merited, and recent grants show that the government is already moving in this direction. Indeed, the federal government recently announced a “Big Data” initiative that provides funds to agencies like the Department of Energy and National Institute of Health to improve data collection and analytics.<sup>307</sup> This program could be expanded to help support the BJS’s similar efforts to assist state and local government data collection and analysis in the criminal justice area.

These and other examples from the criminal history context show that the federal government has an extensive array of tools and experience in the field of incentivizing state and local information collection and sharing. An effort to create a national database of criminal justice information containing limited data such as arrest date, appointment of counsel, and pretrial hearings could build from the lessons learned from efforts in the criminal history context. The fact that states already have a nonprofit organization that works with the federal government and negotiates the uniformity and sharing of state and local criminal data<sup>308</sup> will also greatly aid this effort, preventing the need to form new organizations for this task. Nonetheless, a broad-based reform of the way criminal justice information is collected and shared is a significant undertaking, and, as with any proposal for a new government program, cost is a concern. So, too, is privacy: disseminating unredacted data could be extremely damaging to both convicted and unconvicted defendants. These objections are considered in the following Part.

---

<sup>303</sup> *State Justice Statistics Program*, *supra* note 188.

<sup>304</sup> *Criminal Justice Data Improvement Program*, *supra* note 187.

<sup>305</sup> *See, e.g., State Profiles*, *supra* note 189 (showing funding for a project in Hawaii to automate certain information about supervised adult parolees and felony probationers for easier compilation and statistical analysis).

<sup>306</sup> *See* BUREAU OF JUSTICE STATISTICS, *supra* note 190.

<sup>307</sup> Press Release, Exec. Office of the President, Obama Administration Unveils “Big Data” Initiative: Announces \$200 Million in New R&D Investments (Mar. 29, 2012), [https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/big\\_data\\_press\\_release\\_final\\_2.pdf](https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/big_data_press_release_final_2.pdf) [<https://perma.cc/BNQ8-HXZU>].

<sup>308</sup> *About Us*, SEARCH, <http://www.search.org/about-search/> [<https://perma.cc/82LD-4TAB>] (“As a national nonprofit organization of the States, SEARCH is the premier resource for collecting, sharing, and analyzing innovative and timely knowledge, information, best practices, services and solutions for justice information sharing.”).

## V. OBJECTIONS

Beyond state and local governments' predictable resistance to further federal demands for data provision, a variety of objections to the federal criminal justice database are likely. The strongest resistance will likely arise from concerns about the resources required of states and local governments that already struggle to make ends meet. Further, similar to privacy concerns that have already arisen from private and public entities' growing use of criminal histories, there will be objections to further collecting, categorizing, and disseminating information about pretrial decisions, case disposition, sentencing, and other aspects of criminal defendants' cases. Neither of these objections, however, appears to outweigh the potentially extensive benefits of improving access to this criminal justice information.

### A. *Costs*

Because state and local governments, federal agencies, and courts already collect some of the information that would be included in a criminal justice database, updating existing systems to link with a federal database would not be as expensive as creating a new system from whole cloth. Many state and local governments produce and record information such as the date on which counsel was appointed, the date of the pretrial hearing, and case disposition. Some also collect basic demographic data for each defendant, such as race and income.<sup>309</sup> Although data inputs would have to be made uniform, and in some cases governments would have to change and expand the type of data that they record for each defendant, this need not be a herculean task, at least for jurisdictions already electronically inputting data.

Indeed, the database proposed in this Article could, and likely should, be implemented incrementally. The federal database manager could first collaborate with states like New York that already collect uniform information from county and municipal courts.<sup>310</sup> The manager could work with the states to update these state systems and link them with the national database. Further, the manager and initial participating states could identify the most challenging aspects of this effort, such as rates of data input errors—particularly for new categories of information that state and local governments did not previously record—before expanding the database to other jurisdictions. These lessons learned could make expansion to additional states more efficient. The database would, of course, be most useful if all jurisdictions participated. But the dearth of existing uniform, accessible data is so stark that any improvements—even the initial inclusion of just two states within the database, for example—could support important analytical efforts to identify needed criminal justice reforms or support existing reform efforts.

---

<sup>309</sup> See *supra* notes 145–48 and accompanying text.

<sup>310</sup> See *supra* notes 155–59 and accompanying text.

## B. *Privacy Intrusions*

An additional objection to the uniform, consistent, and widespread availability of criminal justice information is privacy—a concern already raised by the many courts that have rules prohibiting electronic access to data or bulk searches of that data.<sup>311</sup> These courts worry that full electronic access will allow identity theft; encourage more crime against particular, targeted individuals; and compromise individuals' privacy,<sup>312</sup> among other problems.

States like Idaho have recognized that these concerns can be relatively easily solved by omitting certain information from public records.<sup>313</sup> This effort is somewhat costly, as it requires removal of data that is included within case summaries and dockets. For example, a committee appointed to examine public access to court records in Maryland noted that “court records consist of case docket sheets, which contain information identifying the parties (name, address, and in criminal cases date of birth, height, weight, sex and race) and describing case events (such as filing and disposition).”<sup>314</sup> The courts must continue to collect this information for their own purposes, but automated programs that eliminated personal information before making the data electronically accessible to the public would reduce the cost of the data redaction or removal effort. Further, just as states like New York have an automated program to search for errors and anomalies in electronic records,<sup>315</sup> as a back up to automated removal of personal information courts could run a similar program to double check records for privacy purposes.

Thus, while there are legitimate concerns about broadly distributing basic case summary and docket data from criminal cases, there are reasonable responses to these concerns. Just as technology makes it possible to more easily record and disseminate data that will inform criminal justice reform efforts in important ways, technology also enables the exclusion of certain data that should not be distributed to the public.<sup>316</sup>

## VI. CONCLUSION

Information alone cannot promise meaningful change in the criminal justice system. Even the most thorough data will not guarantee a successful court challenge, particularly in areas like Equal Protection challenges where discriminatory intent must be shown in addition to discriminatory effects.<sup>317</sup> Nor will better information necessarily cause nonprofit and governmental

---

<sup>311</sup> See *supra* notes 142–43 and accompanying text.

<sup>312</sup> See *supra* notes 142–43 and accompanying text.

<sup>313</sup> See, e.g., *Idaho Supreme Court Data Repository*, *supra* note 46.

<sup>314</sup> See MD. COURTS, *supra* note 139, at 11.

<sup>315</sup> See *supra* note 164 and accompanying text.

<sup>316</sup> Cf. Gibson, *supra* note 232, at 166 (“[D]igital technology can both liberate information and enclose it.”).

<sup>317</sup> See *Washington v. Davis*, 426 U.S. 229, 240 (1976).

groups to target the areas of the criminal justice system that seem to demand the most attention. These groups respond to a variety of signals, including personal experiences of their staff, anecdotal evidence from the areas of the country in which these groups typically work, and the priorities of the largest donors—preferences that might also be shaped by personal experience and anecdotes.<sup>318</sup> These perceptions are durable, and empirical evidence pointing to areas where the greatest injustices occur will not immediately cause foundations and other nonprofit groups working in the criminal justice area to change course.<sup>319</sup> Further, statistical data pointing to jurisdictions that systematically, egregiously violate the basic rights of criminal defendants by failing to appoint counsel for indigent defendants or holding defendants pretrial for extended periods of time is not guaranteed to sway policymakers. Local officials, in particular, might be more likely to dig in their heels and become defensive in response to proof that certain aspects of their criminal justice system lags behind others.

Despite the limits of empirical data, some recent examples show the potential for broadly accessible uniform data to effect substantial change. The imprisonment of millions of individuals under harsh crack cocaine sentences led to a gradual recognition of largely unsupported variations in sentencing and to sentencing reforms.<sup>320</sup> Even more recently, the rise of digital videos produced from handheld smart phones has called attention to injustices during the arrest movement and has helped spur the Black Lives Matter movement.<sup>321</sup> And where groups have been able to collect detailed information about routine problems in the criminal justice system by sending staff to regularly sit in courtrooms at pretrial hearings,<sup>322</sup> or by conducting extensive surveys of

---

<sup>318</sup>Cf. Mark Seidenfeld, *Why Agencies Act: A Reassessment of the Ossification Critique of Judicial Review*, 70 OHIO ST. L.J. 251, 268–86 (2009) (exploring agency officials' incentives).

<sup>319</sup>See, e.g., Hannah J. Wiseman, *Regulatory Islands*, 89 N.Y.U. L. REV. 1661, 1720–21 (2014) (noting how nonprofit groups respond to a variety of incentives, including the preferences of donors in the environmental context).

<sup>320</sup>See, e.g., Fair Sentencing Act of 2010, Pub. L. No. 111-220, 124 Stat. 2372 (codified in scattered sections of 21 and 28 U.S.C.); see also David A. Sklansky, *Cocaine, Race, and Equal Protection*, 47 STAN. L. REV. 1283, 1285–97 (1995) (describing the influence of crack cocaine in initially encouraging mandatory minimum sentences, including for narcotics trafficking, that contributed to large racial disparities in the criminal justice system).

<sup>321</sup>Mario L. Barnes, *Foreword: Criminal Justice for Those (Still) at the Margins—Addressing Hidden Forms of Bias and the Politics of Which Lives Matter*, 5 U.C. IRVINE L. REV. 711, 713–14 (2015).

<sup>322</sup>See, e.g., *Heckman v. Williamson County*, 369 S.W.3d 137, 160 (Tex. 2012) (failing to appoint counsel for indigent defendants among other potential problems). Staff of the Texas Fair Defense Project, which initiated the case, attended Williamson County hearings in which the defendants were apprised of the charges against them and they documented the defendants' requests for counsel and the routine denials of counsel at the hearings. E-mail from Harry Williams to author, *supra* note 114. Staff were able to access

practices within limited geographic areas,<sup>323</sup> they have tended to succeed in obtaining change through the courts.<sup>324</sup>

Big data in the criminal justice context will likely have the most potential when used to identify problems in the criminal justice system that have not yet been fully documented. For example, although the right to counsel for indigent defendants was solidified in 1963, inadequate access to counsel remains a massive problem.<sup>325</sup> And although this problem has been documented by numerous studies providing estimates of underfunding and other representation problems, solid, empirical data could provide information on the localities and states that pose the largest problem and should be the targets of reform. Similar data is also critical to revealing systemic problems, and jurisdiction-specific failures, in the contexts of bail and debtors' prisons, among other contexts.

Given the realities of our criminal justice system—consisting of thousands of largely autonomous and heterogeneous jurisdictions—bringing criminal justice data collection into the twenty-first century will not be an easy task, but neither is it an insuperably difficult one. But given the potential payoffs from even incremental steps towards that goal, both in facilitating broad-level research and identifying problems at the local level, it is one worth pursuing.

---

online data to determine when counsel were not appointed but were not able to tell from the data why counsel were not appointed. *Id.*

<sup>323</sup> For example, the Equal Justice Under Law project closely investigates bail practices within certain counties and then files class action challenges against these practices. The group has succeed in several of these challenges. *See supra* note 18.

<sup>324</sup> *See supra* notes 322–23.

<sup>325</sup> *See supra* notes 21, 24, 112 and accompanying text.

