THE COLUMBIA
SCIENCE & TECHNOLOGY
LAW REVIEW

VOLUME XXVI STLR.ORG FALL 2024

ARTICLE

DATA INFRASTRUCTURE AS COURT ARCHITECTURE

Kat Albrecht*

Whether courts like it or not, digital legal data has become an important part of both litigation and justice administration. Constitutionally protected as public records, court data and court-adjacent data must be made transparent and accessible to the general public. However, alongside considerations of how to make court data accessible externally, so must we also consider how court data is situated internally within courts. Conceptualizing data infrastructure as court architecture reframes the importance of court data to better align with its current utility in courts, while privileging the very real structural issues that courts must contend with to ensure the continued health of data systems. This Article considers the usefulness of data in the current terrain of law and justice, evaluates the ecosystem of data products currently at play in all levels of courts, and offers concrete pathways to data infrastructure development through Open Knowledge Networks.

* Kat Albrecht is an Assistant Professor at Georgia State University in the Andrew Young School of Policy Studies. She received her JD and PhD degrees from Northwestern University and the Northwestern Pritzker School of Law. At the time of this writing, she was also a fellow with Georgetown University Law Center's Judicial Innovation Fellowship embedded within the Hamilton County, Tennessee General Sessions Courts and an MFA student at the University of Georgia. The author is grateful to the employees of the Hamilton County, Tennessee, General Sessions courts, from whom she learned a great deal about what it means to be a court within a community. She is also appreciative of the feedback and support from her cohort of Georgetown Judicial Innovation Fellows Verenice Ramirez and Emily Lippolis.

I. INTRODUCTION

A large volume of work, including my own, calls for the expansion of accessible court data as imperative to the project of facilitating access to justice via access to data. Calls for public data remain of the utmost importance not only for vindicating the rights of individuals but also for auditing the functions of the legal system at scale.[1] While scholars, practitioners, and policymakers advocate for data availability, sometimes including building external tools and generating solutions around data voids, there is a dearth of solutions-oriented work acknowledging the specific and tangible data infrastructure challenges that need to be alleviated *within* courts themselves. These infrastructure pain points are sharply felt in state and local courts which have the distinction of resolving a vast majority of legal issues while also facing persistent underfunding that functionally blocks innovation.[2]

In addition to enumerating data infrastructure issues that local courts face, this Article expressly considers how the same data infrastructure required to optimally deliver data to the public would also support the efficient internal functions of courts. Conceptualizing data infrastructure as court architecture is of increasing importance as data eclipses its status as archival ephemera to take on a more central role in both legal practice and the administration of justice. As courts almost inevitably make the transition towards digital recordkeeping and connected data systems, there are also new investment areas that courts must consider in order to keep data systems healthy and secure in ways that support justice administration.

---

[1] *See generally* David L. Schwartz et al., *The SCALES Project: Making Federal Court Records Free*, 119 NW. U. L. REV. 23, 27-28 (2024) (explaining the legal right to access court records and the utility of court records in both auditing the justice system at scale and providing information back to the general public).

[2] See Ian Weinstein-Tull, *Local Judges and Local Government*, 106 VA. L. REV. 1031, 1031-1034 (2020).

I lay out this Article in four parts where Part II considers the utility of legal data and contemporary demands of data systems, Part III focuses on data infrastructure and case management in state and local courts, Part IV introduces and evaluates knowledge graph architecture as a potential blueprint for re-imagining court data infrastructure, and Part V concludes with best practices recommendations at different levels of resource availability.

II. THE UTILITY OF DATA AND THE DEMANDS OF DATA SYSTEMS

We must necessarily begin by evaluating whether data is useful enough to justify the time, money, and complexity that reinvigorated and redesigned data infrastructure will demand. I present an analysis of data utility that considers what legal imperatives already exist demanding public access to data, how the current unideal distribution of court data shapes legal knowledge, and how new laws and justice administration practices have changed demands for data. I evaluate these imperatives in order, beginning with access to data as access to justice, following with an analysis of how constrained data constrains knowledge, and finally by considering the emergence of 'data law' and the pipeline by which administrative data about courts and the legal system has become necessary to vindicate individuals' rights under the law.

A. Access to Data as Access to Justice

Proponents of data transparency as a pathway to justice find strength in the law itself and its protections of public data access. Public right to data is a story that actually begins with the public right to view legal proceedings. In *Richmond Newspapers, Inc. v. Virginia*, the court observed that a closed court serves to build public distrust and takes on an appearance of injustice best remedied through open and observable court proceedings.[3] Not only do courts recognize the rights of the public to attend and observe proceedings, but courts more urgently recognize that their very legitimacy is based on this transparency.[4] Consequently, we have seen the right to attend court proceedings reified in other courts' decisions,[5] and a majority of U.S. courts remain theoretically open for public attendance and observation.

Beyond transparent proceedings, courts have also upheld the right of the public to view court records and documents. In *Nixon v. Warner Communications*, the courts identified this right as a general right that specifically extends to judicial records and documents.[6] This was confirmed and extended in *Associated Press v. U.S. District Court*, which held that courts were required to provide pretrial records

---

[3] Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 571-72 (1980) (citing Offutt v. United States, 348 U.S. 11, 14 (1954)).

[4] Schwartz et al., supra note 1, at 28.

[5] *See, e.g.*, Globe Newspaper Co. v. Superior Court, 457 U.S. 596, 605-06 (1982).

[6] Nixon v. Warner Communications, 435 U.S. 589, 597 (1978) (holding that "the courts of this country recognize a general right to inspect and copy public records and documents, including judicial records and documents").

under the First Amendment.[7] But, that is not to say that all records are automatically public or even meaningfully public. Instead, courts have specified that they maintain substantial amounts of discretion in how much data they actually make publicly accessible.[8] Additionally, scholars have found a universe of difference between theoretical public access and actual public access, where court data is de jure public but de facto non-public.[9]

Even Public Access to Case Electronic Records ("PACER"), a service provided by the U.S. government ostensibly designed to make all federal court records publicly accessible, is not *de facto* public. PACER suffers from a fee structure requiring ten cents per page for searching and downloading documents, making it functionally unusable at scale even for individuals who understand how to navigate its archaic search structure.[10] Pulling docket reports, which are just a small piece of the available records, for federal court cases filed in 2016 costs over $100,000.[11] The price tag and expert knowledge that accompanies a viable search for public records casts these records as not realistically public. Beyond demonstrating how PACER is ineffective at its named charter, the PACER example is also a cautionary tale of how theoretically useful tools can be hamstrung by their specific type of deployment and funding structures.

These problems become even more acute when resolved at the local or state level, where existing administrative data largely comes from archaic record keeping systems that vary in how much information they make available and in what form. Where court data does exist, it is often collected and maintained in decentralized, jurisdictionally siloed ways that make reconstituting the broader picture of the local justice systems even at the state level exceptionally difficult.[12] Where some courts have a data portal of some varying form of accessibility, some have nothing at all.[13]

In an investigation of data records requests at the county level, Albrecht and Filip found that requirements for records requests, if such a procedure could be

---

[7] Associated Press v. U.S. District Court, 705 F.2d 1143, 1147 (9th Cir. 1983).

[8] Ronald D. May, *Public Access to Civil Court Records: A Common Law Approach*, 39 VANDERBILT L. REV. 1465, 1469 (1986); *See also* Peter W. Martin, *Online Access to Court Records—From Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 855, 860-65 (2008) (detailing the origins of PACER).

[9] Kat Albrecht & Kaitlyn Filip, *Public Records Aren't Public: Systemic Barriers to Measuring Court Functioning & Equity*, 113 J. CRIM. L. & CRIMINOLOGY 1, 6-7 (2023).

[10] John L. Moreland, *Is Open Access Equal Access? PACER User Fees and Public Access to Court Information*, 49 DTTP: DOCUMENTS TO THE PEOPLE 42, 43-44 (2021).

[11] Adam R. Pah et al., *How to Build a More Open Justice System*, 369 SCIENCE 134, 136 (2020).

[12] *See* DAVID W. NEUBAUER & HENRY R. FRATELLA, AMERICA'S COURTS AND THE CRIMINAL JUSTICE SYSTEM 102-103 (13th ed. 2019).

[13] See *Methodology*, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/methodology/ [https://perma.cc/4ZMZ-NK25] (last visited Nov. 13, 2024). For examples, see Nebraska State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/ne/ [https://perma.cc/97ZS-WNNT] (last visited Nov. 13, 2024); Nevada State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/nv/ [https://perma.cc/6PDA-R33X] (last visited Nov. 13, 2024).

located at all, varied substantially by county sometimes even requiring the specific case date, the name of the judge, and the number of pages in the document, which they write, "Functionally means that a member of the public must be a party to that case, know a party to that case, or possess expert-level knowledge to find the relevant information for a given case. This again constitutes a barrier to true public access."[14] This represents a system that is technically public but is not meaningfully public, since using that system would require knowledge and resources not generally known to the public.

While there is a spectrum of court data access from non-existent to government-systematized, even the most ideal mass resolutions of court data infrastructure do not adequately and meaningfully uphold rights to public access at the national-level.[15] As this Article will go on to consider, this terrain of access becomes increasingly complex as private vendors enter the court data arena, creating new types of pain points and harms as they try to profit off a need for data system efficiency.[16]

B. How Data Availability Constrains Knowledge

Other avenues do exist to dispense records related to the justice system, including mechanisms like public information request laws. However, the way that data is created, maintained, and distributed puts significant constraints on what is actually knowable about the legal system. The most known of the public record request laws, the Freedom of Information Act ("FOIA"), theoretically provides members of the public a means to obtain public information about federal government agencies and their activities.[17] However, in practice, validly submitted FOIA requests often fail to provide responsive information and agencies regularly fail to observe disclosure deadlines.[18] FOIA lawsuits have increased in recent years, with analysis showing that an increasing volume of lawsuits derives not from substantive issues, but from agency non-response.[19]

---

[14] Albrecht & Filip, *supra* note 9, at 31.

[15] This Article will go on to identify specific places across the U.S. where there is innovation in data access and transparency, often holding them up as positive examples of the viability of data transparency and data sharing, but I argue here that the constitutional right to data has not been vindicated across the U.S. as a whole.

[16] Jason Tashea, *A Human Rights Approach to Justice Technology Procurement*, GEO. L. TECH. REV. (Sept. 2023), https://georgetownlawtechreview.org/a-human-rights-approach-to-justice-technology-procurement/GLTR-09-2023/ [https://perma.cc/RM6V-AUJK].

[17] 5 U.S.C. § 552. Note that while FOIA is federal, it also serves as a blueprint for state-level information request acts.

[18] Paul Kiel, *Delayed, Denied, Dismissed: Failures on the FOIA Front*, PROPUBLICA (May 26, 2010), https://www.propublica.org/article/delayed-denied-dismissed-failures-on-the-foia-front [https://perma.cc/395R-GDT3].

[19] *FOIA Suits Rise Because Agencies Don't Respond Even As Requesters Wait Longer To File Suit*, THE FOIA PROJECT (Dec. 15, 2019), https://foiaproject.org/2019/12/15/foia-suits-rise-because-agencies-dont-respond-even-as-requesters-wait-longer-to-file-suit/ [https://perma.cc/2WGG-Q5FX].

As part of a recent project on the national picture of felony murder cases, journalist Sarah Stillman and her team sent FOIA requests to U.S. States requesting information about the number of individuals convicted or currently incarcerated with felony murder charges.[20] The request also asked for demographic information about individuals like age, sex, and race. Following a round of requests and appeals, Stillman's team was still unable to obtain analyzable data for a majority of U.S. States. Ultimately, by incorporating data extracts held by academics and advocacy groups, Stillman's team was able to reach a total of eighteen states where they had usable data for analysis.[21]

Along the way, Stillman's project revealed various failure points in the FOIA process and demonstrated how the way we store court data functionally limits knowledge. Several states simply did not respond or abruptly stopped responding before completing the request.[22] Other states cited a residency requirement meaning they did not have to produce data at all.[23] Some requests were met with a response stating that the state did not have data responsive to the request.[24] A large number of requests were returned with data that did not actually identify felony murder as stipulated in the request. Sometimes the responses consisted only of links to a public interactive data dashboard that did not allow Stillman's team to specifically identify felony murder at all.[25] These discouraging results actually represent a

---

[20] *About*, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/about/ [https://perma.cc/L38X-D9JX]. The investigation excluded states who do not have felony murder laws on the books, namely Hawaii and Kentucky.

[21] *State Data*, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/ [https://perma.cc/S5GS-AYHR] (last visited Nov. 13, 2024).

[22] For examples, see Maryland State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/md/ [https://perma.cc/6Y8M-5Z8D] (last visited Nov. 13, 2024); Utah State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/ut/ [https://perma.cc/YV8Z-W88X] (last visited Nov. 13, 2024). In the case of Utah, Stillman's team exchanged ten emails with an agent of the state before the state employee's email was abruptly cut off entirely rendering further communications undeliverable. *Id.*

[23] For examples, see Tennessee State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/tn/ [https://perma.cc/JF3J-9KRX] (last visited Nov. 13, 2024); Alabama State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/al/ [https://perma.cc/ZTY2-SRG6] (last visited Nov. 13, 2024); Arkansas State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/ar/ [https://perma.cc/7XAW-UUPW] (last visited Nov. 13, 2024).

[24] For example, from Virginia, Stillman's team was told, "Please be advised the Virginia Department of Corrections (VADOC) does not have any records responsive to your request because the records do not exist. Pursuant to § 2.2-3704(D) of the Code of Virginia, 'no public body shall be required to create a new record if the record does not already exist.'" Virginia State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/va/ [https://perma.cc/LF6C-ECNA] (last visited Nov. 13, 2024).

[25] For examples, see Nebraska State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/ne/ [https://perma.cc/97ZS-WNNT] (last visited Nov. 13, 2024); Nevada State Data, THE FELONY MURDER REPORTING PROJECT,

multi-faceted problem. For some states, the problem really was that the records exist somewhere within the state but were not, for whatever reason, produced. But in other states, they do not specifically identify felony murder in their general administrative data—greatly lowering the chance of receiving a fully responsive dataset to *any* request involving felony murder at scale. The lack of nuanced information holdings and distribution suggests an insufficient internal structure for facilitating data transfer outward from the courts and other justice administrative units.

If every state had these problems, it would be tempting to say that it just is not possible to create high-quality data for public consumption. However, in 18 states, Stillman's team was able to gather analyzable data,[26] indicating that sufficiently detailed data production is perfectly possible. Four states in the United States use one specific legal statute to identify felony murder,[27] a level of uniqueness that makes it easier to locate felony murder within court data. However, even when the codification of felony murder was more complex, some states were still able to produce actionable data.[28] More serious difficulty resulted when the states just did not separate felony murder from other types of murder and lacked developed internal systems to track that information clearly.[29] Most importantly, where Stillman's team was unable to procure analyzable data, there is no analysis. This means that a detailed analysis of the national picture of felony murder is not really a national-level analysis. Instead, it is an analysis of jurisdictions with sufficiently developed systems to produce data, leaving unanswered questions about equity and bias in unanalyzed jurisdictions. Beyond the substantive findings from the states that provided data, Stillman's Felony Murder Project also demonstrates how a lack of publicly accessible data leaves knowledge gaps in the national understanding of felony murder, and how internal data infrastructure and record-keeping systems

---

https://felonymurderreporting.org/states/nv/ [https://perma.cc/6PDA-R33X] (last visited Nov. 13, 2024); West Virginia State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/wv/ [https://perma.cc/U2TG-LQC4] (last visited Nov. 13, 2024).

[26] Home Page, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org [https://perma.cc/N5X7-GJTL] (last visited Nov. 13, 2024).

[27] These states are Wisconsin (WIS. STAT. ANN. § 940.03 (2023)), Pennsylvania (18 PA. STAT. AND CONS. STAT. ANN. § 2502 (West 2023)), Missouri (MO. REV. STAT. § 565.021 (2023)), and Illinois (720 ILL. COMP. STAT. ANN. 5/9-1 (2023)).

[28] For examples, see Florida State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/fl/ [https://perma.cc/69BG-RTV6] (last visited Nov. 13, 2024); Mississippi State Data, THE FELONY MURDER REPORTING PROJECT, https://felonymurderreporting.org/states/ms/ [https://perma.cc/2LW9-WNWQ] (last visited Nov. 13, 2024). In Florida, Stillman did not even have to request the data because it was already meaningfully public. Mississippi has a multi-part felony murder statute, and the data received was responsive to the request.

[29] A tempting argument here would assert that maybe separating felony murder from other forms of murder is not undertaken by courts because felony murder is not a unique enough category. However, felony murder differs substantially from other types of murder in requirements for intent, potential distance from the murderous assault, and in punishment.

may render courts unable to easily answer questions about felony murder even internally.

This is potentially very problematic from a data analytics perspective if the states that do not provide data are not missing at random and are instead missing for broader systemic reasons.[30] This problem of systematically missing data happens with courts across the United States. In jurisdictions like Chicago and Philadelphia, where data transparency initiatives from the district attorney's office have made detailed data available, we are able to glean more information about the functions of the legal system in those locations.[31] For example, a search of the National Archive of Criminal Justice Data reveals approximately three times as many studies responsive to Pennsylvania court data compared to Arkansas or North Dakota.[32] These information asymmetries are more dramatic at the state and county levels, where a dearth of data serves to exclude swaths of the United States—and their courts—from analysis at all.[33] In general, larger and more urban areas have more data access and availability, [34] which increases our knowledge of those spaces, but is perhaps damaging to our knowledge of differently situated spaces.

Take for instance, the terrain of studies surrounding crime in the United States. Crime is often erroneously conceptualized as an exclusively urban problem, where even if crime occurs in rural spaces, it is the same as urban crime: just less.[35] However, more recent scholarship has found that is not the case, arguing, "Rural crime and justice are not simply scaled-down versions of urban crime and justice. Rural crime and justice take place in a context that colors the process of how crime develops and how justice is carried out."[36] Consistent with this argument, substantial scholarship has found that urban, rural, and mid-sized communities vary

---

[30] *Missing Data and Multiple Imputation*, COLUM. U. MAILMAN SCH. OF PUB. HEALTH, https://www.publichealth.columbia.edu/research/population-health-methods/missing-data-and-multiple-imputation [https://perma.cc/X8VS-TCAC] (last visited Nov. 13, 2024).

[31] *Research at the DAO*, PHILA. DIST. ATT'Y'S OFF., https://data.philadao.com/Research.html [https://perma.cc/44JP-ZWJ9] (last visited Nov. 13, 2024). *Felony Dashboard*, COOK CNTY. STATE'S ATT'Y, https://www.cookcountystatesattorney.org/about/data-reports [https://perma.cc/C4RP-HRRZ] (last visited Nov. 13, 2024).

[32] This was determined via a search of the National Archive of Criminal Justice Data, NACJD, https://www.icpsr.umich.edu/web/NACJD/search/studies?q= [https://perma.cc/JZQ9-FLGZ] (last visited Nov. 13, 2024) (search for 'state court data'; then refine to the statewide jurisdiction of Pennsylvania, Arkansas, or North Dakota, to exclude general nationwide datasets).

[33] For example, in Sarah Stillman's Felony Murder Reporting Project, states without sufficient quality data were excluded from the final analysis because the data did not allow for statistically valid findings. *See also* Caitlin Glass, Kat Albrecht & Perry Moriearty, *Prosecutorial Data Transparency and Data Justice*, 119 NW. U. L. REV. 1. 193, 193-220 (2024) (finding that responsive, detailed, public data like those offered in Chicago and Philadelphia are still the exception rather than the rule).

[34] Albrecht & Filip, *supra* note 9, at 26 (finding that larger, more urban counties were more likely to have explicit policies to access and acquire data, though references to certain types of data remained sparse).

[35] RALPH A. WEISHEIT ET AL., CRIME AND POLICING IN RURAL AND SMALL-TOWN AMERICA 1, 2 (Waveland Press, 2nd ed. 2005).

[36] *Id.*

on a number of other important vectors that affect crime, including employment, poverty, social capital, and drug overdose.[37] The end result is that we run the risk of generalizing conclusions from jurisdictions who do produce data onto those who do not, despite known differences between these places. This is all to say, some of the things we think we know about legal outcomes and justice administration in the United States are actually things we know for the jurisdictions that create, maintain, and produce data, which may be systematically different from jurisdictions that cannot, will not, or do not.

C. The Rise of Data Law

There are a variety of ways that court data can be useful, including a multitude of uses by the general public, but in the section to follow I focus specifically on data use within courts during active case proceedings.[38] Data is used by courts in a number of ways, with mixed efficacy and ethicality. Here, I briefly consider data in three forms: (1) data related to the validity of scientific techniques; (2) data found to be accurate or compelling, but not legally relevant, and; (3) changes in law explicitly allowing population-level data and broader expert testimony beyond the facts of a currently litigated case. Over the course of these three forms, I argue that there has been a substantial paradigm shift, moving data from the confines of scientific validity checks to using data as a key factor in litigation.

Data in courts has substantially changed in its utility over time, notably shifting from a project of evaluating externally produced evidence to one also required to evaluate evidence produced by courts themselves. Early uses of data generally focused on evaluation of external processes and whether that data should be admitted into courts as evidence. A foray into evidence evaluation in courts finds substantial early literature around expert testimony and the introduction of novel scientific techniques for evidentiary consideration. This era of data use in courts required the court to establish some sort of standard to determine what scientific or data techniques could be used to produce valid legal evidence. This task was made more difficult given that courts not only have to understand the substance of many complex scientific techniques, but they must also decide how to incorporate science

---

[37] Steven Lannoo et al., *Differences in Social Capital Between Urban and Rural Environments*, 34 J. URB. AFF. 373, 382-90 (2012); Paul R. Amato, *Urban-rural Differences in Helping Friends and Family Members*, 56 SOC. PSYCH. Q. 249, 254-59 (1993); Holly Nguyen & Brandy R. Parker, *Assessing the Effectiveness of New York's 911 Good Samaritan Law—Evidence from a Natural Experiment*, 58 INT'L J. DRUG POL'Y 149, 152-154 (2018); Veronica A. Pear et al., *Urban-rural Variation in the Socioeconomic Determinants of Opioid Overdose*, 195 DRUG & ALCOHOL DEPENDENCE 66, 68-72 (2019).

[38] A considerable nascent field has dedicated itself to thinking about how expanded legal data could be useful to the general public and policymakers. *See, e.g.*, Stephanos Bibas, *Transparency and Participation in Criminal Procedure*, 81 N.Y.U. L. REV. 911, 935 (2006) (pointing out that sentencing data is not generally available to the public, so they are unable to take on fully considered opinions about elected prosecutors); Schwartz et al., *supra* note 1, at 28 (presenting a survey of applications for the SCALES Data Explorer and overviewing general insights from legal data).

into law and larger patterns of legal decision making.[39] The *Frye* standard required that a scientific method or technique achieve broad recognition and acceptability within an expert's field study in order for evidence produced by that technique or expert testimony relying on that technique to be admissible in court.[40] The *Daubert* standard usurped *Frye* in many jurisdictions and called on courts to come up with a new way to determine the validity of scientific methods and data that did not functionally disparage newness like *Frye* did. Instead, the *Daubert* standard considered factors like whether or not a method had been tested, if there was peer reviewed research on the method, whether the method was proven reliable, and if there were consistent and valid standards for use of the method.[41] This task has a certain amount of conceptual clarity where it concerns scientific innovations or methods fully separate from the legal system. That is not to say that determining the merit of scientific methods, data, or evidence is easy, but rather that a court evaluating the merits of a fully separate scientific technique does not indict the legal system itself as the creator, maintainer, and producer of that data.

A second distinct consideration of data in courts concerns the dismissal of data that is substantively valuable and reliable as not dispositive in a particular case. In a number of pivotal cases where courts are confronted with data evidence substantively related to the case, they have found that the data and its findings may be accurate and compelling but still not legally relevant. In *McCleskey v. Kemp*, a death penalty case that famously used statistical evidence to argue that the rate of death penalty prosecutions was biased by race where Black defendants were sentenced disproportionately harshly, the court concluded that a robust statistical regression was not sufficient because even though they acknowledged the disparities shown in the data, they ruled the evidence did not demonstrate a constitutional violation in this particular trial.[42] This dismissal of data describing patterns in criminal sentencing effectively undercut the role of data in criminal trials, where the type of proof required to prove individual intentional discrimination in a particular case is often proof that just cannot be produced because the rationale for discretion-based decisions is simply never entered into the record.[43]

Nearly a decade later, *United States v. Armstrong* created data requirements to prove discrimination that functionally make discrimination impossible to prove.[44] The defendants raised a selective prosecution claim based on a showing that, among all cases involving crack cocaine offenses closed by the Central District of

---

[39] *See generally* Harold L. Korn, *Law, Fact, and Science in the Courts,* 66 COLUM. L. REV. 1080, 1093-94 (1966) (providing a nuanced consideration of how science and the imperatives of law are balanced in the application of scientific evidence).

[40] Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923).

[41] *See* Daubert v. Merrell Dow Pharms., Inc., 509 U.S. 579, 595 (1993) (holding that expert opinions are to be admitted "only if the facts or data are 'of a type reasonably relied upon by experts in the particular field in forming or inferences upon the subject'").

[42] McCleskey v. Kemp, 481 U.S. 279, 297-99 (1987).

[43] *See generally* Glass et al., *supra* note 33, at 193.

[44] United States v. Armstrong, 517 U.S. 456, 459 (1996).

California Federal Public Defender's Office in 1991, every single defendant was Black.[45] The court ultimately held that a defendant seeking discovery to advance a selective prosecution claim must demonstrate racially disparate treatment of "similarly situated individuals," which can serve as evidence that prosecutors possessed discriminatory intent.[46] This requirement seems immediately tautological, with scholarly analysis concluding that "the defendant is confronted with *Armstrong*'s cruel catch-22: he must provide evidence of discrimination to obtain discovery about discrimination."[47] A sticky problem emerges when courts are required to evaluate data produced by an imperfect legal system, where the data itself is incomplete or opaque because of how justice system actors fail to create, maintain, and produce sufficiently detailed data. Data about the legal system necessarily comes from legal system actors like police, prosecutors, courts, and corrections. Where that data is systematically unavailable, the catch-22 of *Armstrong* becomes particularly untenable.

The third type of data use described here brings us abruptly to the present day and to the rise of *data law*, where I argue that data takes on a more intentionally central role in litigation. Rather than conceptualizing data law as a site where data will disrupt traditional legal practice, I position data law as a space where data supplements traditional legal practice in that data evidence has itself become a more viable pathway to legal remedies. That is, moving beyond a sort of 'data is becoming normative, so it is used in more cases,' type of argument, I want to instead examine relatively recent law that specifically turns away from the data dismissal set out in *McCleskey* and *Armstrong*, specifically inviting pattern-type data as evidence of discrimination.[48] Passed into law in 2020, the California Racial Justice Act ("CRJA") directs that "[t]he state shall not seek or obtain a criminal conviction or seek, obtain, or impose a sentence on the basis of race, ethnicity, or national origin."[49] There are two key features of the CRJA that are most useful for the

---

[45] *Id.*

[46] *Armstrong*, 517 U.S. at 465.

[47] Alison Siegler & William Admussen, *Discovering Racial Discrimination by the Police*, 115 NW. U. L. REV. 987, 992 (2021).

[48] *See* Young v. Super. Ct., 79 Cal. App. 5th 138, 152-53 (Cal. Dist. Ct. App. 2022).

[49] 2020 CAL. STAT. 93, ch. 317 § 3.5 (See relevant language from the CRJA sections 4 and 5:

> (4) The defendant was charged or convicted of a more serious offense than defendants of other races, ethnicities, or national origins who commit similar offenses and are similarly situated, and the evidence establishes that the prosecution more frequently sought or obtained convictions for more serious offenses against people who share the defendant's race, ethnicity, or national origin in the county where the convictions were sought or obtained.
>
> (5) (A) A longer or more severe sentence was imposed on the defendant than was imposed on other similarly situated individuals convicted of the same offense, and longer or more severe sentences were more frequently imposed for that offense on people that share the defendant's race, ethnicity, or national origin than on defendants of other races, ethnicities, or national origins in the county where the sentence was imposed.

analysis here: a lower standard of proof and a specific call for pattern evidence. First, in its opening motions, the CRJA aspires to a standard of proof that is a conceptual match for preliminary data evidence. The standard of proof language reads: "The defendant produces facts that, if true, establish that there is a substantial likelihood that a violation of subdivision (a) occurred. For purposes of this section, a 'substantial likelihood' requires more than a mere possibility, but less than a standard of more likely than not."[50]

Scholars interpreting the CRJA in light of the California Penal Code note that descriptive statistical data evidence is sufficient at a "more than a mere possibility level," asserting: "While [descriptive] statistics alone cannot show whether purposeful racial discrimination occurred in an individual case, no such showing is required under the CRJA. [Such] statistics *can* establish 'more than a mere possibility' of racial bias. . . ."[51] The combination of these features appears substantially more hospitable for data as evidence in courts than the previous *McCleskey* era. However, in practice, a defined role for system-produced data in the courtroom necessarily demands that the internal systems of courts create, maintain, and produce accurate data. This pipeline for usable and sufficient data does not currently exist.

The first Racial Justice Act hearing in San Francisco County, California, tested this data pipeline with discouraging results. In *People v. Fantasy Decuir*,[52] an analysis of racial disparity in special circumstance enhancements led to an in-court presentation of multiple statistical regression models focused on similarly situated comparisons between the defendant Decuir and other victim-defendant dyads.[53] The experts in the case presented a series of models comparing principal charges,

---

> (B) A longer or more severe sentence was imposed on the defendant than was imposed on other similarly situated individuals convicted of the same offense, and longer or more severe sentences were more frequently imposed for the same offense on defendants in cases with victims of one race, ethnicity, or national origin than in cases with victims of other races, ethnicities, or national origins, in the county where the sentence was imposed).

[50] 2020 CAL. STAT. 93, ch. 317 §3.5(h).

[51] Brief of Amici Curiae Fred T. Korematsu Center for Law and Equality et. al in Support of Petitioner, at 18, Mosby v. Super. Court, 99 Cal. App. 5th 106 (2024) (No. RIF1604905) (citing CAL. PENAL CODE § 745(h) (West 2023)) (emphasis in original).

[52] The author of this Article was one of the computational science experts on this case who participated in analyzing discovery, building the data, analyzing the data, preparing filings for the court, and testifying at trial.

[53] *See* Alex Carney & Kat Albrecht, *Analysis of Racial Differences in Life Without the Opportunity of Parole Charges in San Francisco County*, filed as Exhibit A in Support of Motion to Dismiss the Special Circumstances Allegation Pursuant to California Racial Justice Act, People v. Fantasy Decuir, No. CRI-17011544 (Cal. Super. Ct. San Francisco Cnty. filed Oct. 8, 2017); *See* Petition for Writ of Mandate, Prohibition, or Other Extraordinary Relief at 37-41, Decuir v. Super. Court, No. A170881 (Cal. Ct. App. July 12, 2024). The analysis in this case followed in the tradition of McCleskey. S*ee, e.g.*, David C. Baldus et al., *Arbitrariness and Discrimination in the Administration of the Death Penalty: A Challenge to State Supreme Courts*, 15 STETSON L. REV. 133, 195 (1986); David Baldus et. al, *Death by Stereotype: Constitutional Challenges from California's Failure to (Again) Narrow Death Eligibility*, 16 J. EMPIRICAL L. STUD., 1394, 1442 (2019).

particular enhancements, and race and gender of victims and defendants.[54] The judge was unconvinced, pointing to a small handful of factual errors in the data and critiquing the validity of the regression models on the grounds that there were unmeasured factors. However, the defense team was left to architect a dataset themselves, where the prosecutor's office did not provide data but were later able to critique that data.[55] The judge cited the data produced in *McCleskey v. Kemp*, arguing that the data work in Decuir needed to be as detailed as the data work undergirding multiple scientific articles that ultimately was not accepted in court.[56] The court did not provide a pathway to obtaining the requested unmeasurable factors. This allusion to a nebulous set of unmeasurable factors without clear data retrieval activities is not unique to Decuir. Courts have asked for data held explicitly by prosecutors or subject to substantial interpretation including the relative strength, availability, and admissibility of the evidence, the viability of potential defenses, the wishes of the victim and their next of kin, and the heinousness of the offenses to be used by defense teams who have no actionable pipeline to creating or obtaining said metrics.[57]

Discovery is not a simple remedy to data retrieval. Even through the process of discovery, a defunct internal system struggles to respond completely to data requests in a timely way if that data exists within the internal system at all. However, the known limitations of internal recordkeeping and data production is not a justifiable release from the requirements of the law, and it is even less justifiable to shift the burden of systemic data failure onto an individual criminal defendant who is pursuing explicitly enumerated remedies via legal data. Moreover, in some data-relevant cases like appeals cases, there is no entitlement to discovery at all. This leaves the current state of data law in a difficult spot where the law explicitly invites the use of data that court data systems are not prepared to create, maintain, or produce.

III. DATA IN LOCAL COURTS

In this Article, I discuss the utility of data within courts for the purposes of both litigation and justice administration. The examples previously described focus on what happens to data within courts when it is used in litigation and when it leaves

---

[54] *See* Alex Carney & Kat Albrecht, *Analysis of Racial Differences in Life Without the Opportunity of Parole Charges in San Francisco County*, filed as Exhibit A in Support of Motion to Dismiss the Special Circumstances Allegation Pursuant to California Racial Justice Act, *People v. Fantasy Decuir*, No. CRI-17011544 (Cal. Super. Ct. San Francisco Cnty. filed Oct. 8, 2017); *See* Petition for Writ of Mandate, Prohibition, or Other Extraordinary Relief at 37-41, Decuir v. Super. Court, No. A170881 (Cal. Ct. App. July 12, 2024).

[55] The response revealed that remediating the alleged errors did not change the analysis due to statistical safeguards and confidence intervals accounting for margins of error in the experts' original analysis.

[56] *See* Petition for Writ of Mandate, Prohibition, or Other Extraordinary Relief at 37-41, Decuir v. Super. Court, No. A170881 (Cal. Ct. App. July 12, 2024).

[57] Brief of Amici Curiae Fred T. Korematsu Center for Law and Equality et. al in Support of Petitioner, at 27, Mosby v. Super. Court, 99 Cal. App. 5th 106 (2024) (No. RIF1604905) (quoting Prosecution's Return to Petition for a Writ of Mandate at 17-18).

courts via public records, but perhaps the most critical uses of data come before either of these events. This pivotal data moment happens in the internal data systems of courts and connected actors where they collect not just what becomes externally used data but also the internally used data that allows everyday operations of the court to be carried out. In Part III, I take up an analysis of this internal world, with specific attention paid to state and local courts. I begin with a series of observations from my own fieldwork before turning to a discussion of case management systems and vendors versus internally developed data products. I conclude this section with a consideration of the current state of innovations and challenges with unifying court data systems.

A. Insights from the Field

Between 2023 and 2024, I spent eleven months embedded within the General Sessions Courts in Hamilton County, Tennessee.[58] Hamilton County is home to approximately 366,207 people and the City of Chattanooga.[59] As of 2023, the median annual income is higher than the state average at $73,611 per household, with 12.6% of the population living below the poverty line.[60] Approximately 70% of Hamilton County residents are white, approximately 17.5% are Black or African American, and approximately 7% are Hispanic or Latino.[61] My charge and charter was to spend the better part of a year analyzing the county's data culture, data systems, and universe of data products. Hamilton County is distinct from similar counties in that it expressly sought out data innovation and systems improvement via partnership with the Judicial Innovation Fellowship program. In this way, I expected Hamilton County to be more open and interested in their data pipelines than courts might be on average.

My first experience in the county court was that of an outsider. I arrived at the court at 600 Market Street at 8:20 a.m., early for a civilian, but a little late for a county court worker. It was not the district court down the street with its stately interiors and regal facades, it was a much more perfunctory building with courtrooms that swelled with noise and the local populace resolving the types of lower-level cases adjudicated en masse in county courts. Over the course of the almost-year, I followed the journey of data through the system with my own observations and interviews. I interviewed and observed over fifty county employees and offices, all focused on their interactions with the data systems. My

---

[58] These observations stem from a related project undertaken as part of the Judicial Innovation Fellowship (JIF) at Georgetown University. For more information about the JIF program and the contours of the project, *see Hamilton County & Judicial Innovation Fellowship Statement of Work*, JUDICIAL INNOVATION FELLOWSHIP (July 6, 2023), https://docs.google.com/document/d/1F-vVUWN5k4iq0o09wgpE7PVUBH6NDZS_reD-UC_VUZ8/edit#heading=h.y18tcbdx4oy1 [https://perma.cc/F3RX-VNM3].

[59] *Hamilton County, Tennessee*, U.S. CENSUS BUREAU, https://data.census.gov/profile/Hamilton_County,_Tennessee?g=050XX00US47065 [https://perma.cc/ZES6-E5LN] (last visited Oct. 4, 2024).

[60] *Id.*

[61] *See id.*

work took me from clerks' offices to the jail, into courtrooms, through conversations with judges, then to visits with community partners and legal aid, out to various municipal courts and police departments, to politicians' offices, and even beyond the borders of Tennessee.[62]

When we think about data innovation and management in courts in general, the literature is resoundingly grim, with courts described as lacking innovation, hamstrung by archaic infrastructure, and dealing with meager budgets—all problems concentrated and exacerbated in underfunded local courts.[63] However, what I found in Hamilton County was an impressive array of locally driven innovations specifically attempting to alleviate real system frustrations. This is not to say that some of the data issues I describe elsewhere in this Article are not also present in Hamilton County, but rather that the employees of Hamilton County have undertaken substantial uncredited data and digital innovation work that have allowed the courts to transition into the modern era. Among these efforts are things that employees in Hamilton County do not necessarily consider hugely innovative but are actually quite advanced compared to the situations some courts find themselves in.

For example, when an arrest enters the Clerk's Office through the so-titled Annex, the arresting officer live-connects via video call with the clerk, who then confirms and initiates the record into the county internal data system.[64] The arrest documents also immediately go into imaging.[65] This pipeline, which I observed working smoothly, represents more than just a data system initiation. Instead, it represents continuity and collaboration between separate agencies to produce a unified data record. This is a pattern of success with the internally developed data system called the Criminal Justice User System ("CJUS"), which is more than a court records database and instead functions as a Case Management System ("CMS") featuring intentional data linkages.

As a criminal case continues through the Hamilton County court system, the internal case management system remains a central feature of justice administration. Hamilton County's system, CJUS, is at the center of this universe. CJUS began as a one-man show, developed by the long-time head of county IT in the early 2000s.[66] CJUS started as a fairly simple data table, but has since expanded to accommodate increasingly complex needs. The creator of CJUS described it to

---

[62] Kat Albrecht, Court Data Innovation in Hamilton County 13, (2024) (unpublished report for Georgetown Law Judicial Innovation Fellowship) (on file with author).

[63] Nancy Scola, "*Courts 'Choose' to Lag Behind on Tech Says Chief Justice Roberts*," WASH. POST (2015); Quintin Johnstone, *New York State Courts: Their Structure, Administration and Reform Possibilities*, 43 N.Y.L. SCH. L. REV. (1999). Larry Berkson, *Unified Court Systems: A Ranking of the States'*, 3 JUST. SYS. J. (1977); Jennifer A. Tallon, Olivia Dana, & Elise Jensen, *The Contradictions of Violence: How Prosecutors Think About the Biggest Challenge to Reform*, CTR. FOR CT. INNOVATION (2022), https://www.innovatingjustice.org/sites/default/files/media/document/2022/Prosecutors_Violence_101122.pdf [https://perma.cc/L69V-U2VK].

[64] *See* Albrecht, *supra* note 62, at 13.

[65] *Id.* at 13.

[66] *Id.* at 11.

me as "adding additions to a house over and over" to make it responsive to new data feature requests and new vendor products that need to interface with the system.[67]

While court clerks use CJUS extensively, checking information and scheduling hearings live in court, other actors enter data into CJUS through a system of unit-specific data portals that are later connected together. For instance, at the jail, deputies must also enter data into CJUS to help case records progress.[68] CJUS does a lot of automatic system calculations—for fines and fees payable, the amount of time an individual has sat in custody between case milestones, information about bond, and more—that actively facilitate justice administration within the county. It also links cases together, presenting a fuller rendering of an individual's history with the county court system. Despite being the development and support team for CJUS, the county is far from the only group benefitting from CJUS. Most of the major municipal courts also use CJUS or are transitioning to CJUS, citing the benefits of being immediately compatible with county systems.[69] Even the crime analysts at the Chattanooga Police Department reported extensive use of CJUS for intelligence investigations.[70]

Given the volume of users and uses of CJUS, you might expect that it is a well-funded and up-to-date system. But it is not. CJUS is supported by four primary developers, significantly more than many small counties across the U.S. have, but still a very small number given the demands of the system.[71] CJUS is internally developed, saving the county lots of money, but is also internally designed and continues to have the visual style of a 2000s-generated product. In some ways, CJUS's dogged adaptability and expansion is its greatest challenge, as it is called to serve more complex purposes without an influx of technological resources.

CJUS is not the only example of intentional system linkages and data innovation in Hamilton County. Hamilton County's juvenile court, which is somewhat but not completely separate from county courts, was the second court in the state of Tennessee to be approved for e-Filing. At the same time as the juvenile court navigated the e-Filing approval process, made substantially more difficult without an external vendor and their pre-made compliance templates, they also set about digitizing over 8.6 million pages of court records. After an 18-month digitization process and the development of a completely free e-Filing system by county IT, the juvenile court now has a fully integrated digital office and digital judicial bench where even self-represented litigants can e-File.[72] The Hamilton County Mayor has also prioritized translating data findings to the general public via the Hamilton Counted Data Initiative. In 2023, the county hired its first ever senior data analyst to aggregate and analyze streams of data on crime, victim

[67] *Id.*
[68] *Id.* at 13-14.
[69] *Id.* at 21-22.
[70] *Id.* at 21.
[71] *Id.* at 20.
[72] *Id.* at 9.

services, substance use, and homelessness for quarterly dissemination to the community.[73]

So far, my observations have primarily described internally developed data products and initiatives, but those are far from the only data products that the County has to contend with. CJUS is not just charged with facilitating its own internal functions, it must also integrate with external vendor procured products like other case management systems, booking technologies, and law enforcement products. This is a tall order because the County IT team responsible for developing these integrated product pipelines are often not included in the procurement process at all for other units.[74] This leaves the 4-person IT team not only responsible for their own features and development, but also adapting to choices made by non-technology stakeholders.

Over the course of my eleven months of observation, recurring challenges with vendor procurement interrupted the ideal of linked data systems across the county. Some of these vendor acquisitions, even if they were met less positively by the county more broadly, received positive reviews from the unit that executed the contract. For example, one municipal court upgraded to a customized version of Incode, a Tyler Technology Product for case management, and reported being generally quite satisfied with the product itself and the provided IT support.[75]

However, the more common outcome was general dissatisfaction with vendor retained products. Beyond dissatisfaction with IT support and daily operations, I identified multiple vendor acquisitions where the new vendor failed to harmonize their offering with the existing data systems in ways that pose significant problems for justice administration and data analysis. For example, the busiest municipal court changed data storage vendors and lost access to data before the vendor change.[76] This poses problems for knowledge continuity in the municipal court but also means that the municipality was excluded from initiatives like Hamilton Counted, which are meant to provide information to the community.

Some of these data infrastructure problems are unanticipated, but they *could be* anticipated if more research and unit-crosstalk undergirded vendor procurement decisions. For example, at the time of my observation, the Chattanooga Police Department was in the midst of negotiating a vendor contract moving to Mark43.[77] The Mark43 contract, which is optioned at nine levels of customization with

---

[73] *Hamilton County Mayor's Office says "Hamilton Counted" reports will bring accountability through data*, LOCAL 3 NEWS (May 11, 2023), https://www.local3news.com/local-news/hamilton-county-mayors-office-says-hamilton-counted-reports-will-bring-accountability-through-data/article_a6d41640-f012-11ed-a647-c794fd7637df.html [https://perma.cc/NM32-F79Y]; Jennifer Baggett, *Hamilton Counted 2023 Year End Report*, HAMILTON COUNTY 3 (2023), https://www.hamiltontn.gov/PDF/Mayor/Media/Hamilton%20Counted%202023%20Year%20End%20Report.pdf [https://perma.cc/X2AS-HCAW].

[74] Albrecht, *supra* note 62, at 23.

[75] *Id*.

[76] *Id.*

[77] *Id.*

increasing costs, had already moved substantially from the originally negotiated level to level 8+, and Mark43 could not confirm that they would be able to conduct data migration from the departments' previous system, a known fail point with other Mark43 contracts.[78] The individuals who initiated the vendor search and initial contract were no longer in those positions, leaving a new team to wade through the bureaucracy and disappointment from multiple directions.[79]

While the context in which I made these observations was unique to Hamilton County, the data infrastructure pain points I identified are not. Rather, the openness of Hamilton County in sharing these pain points suggests that even places pushing at the boundaries of data infrastructure and innovation are still experiencing infrastructure challenges born from a lack of future-thinking, system-level data procurement and data development processes. Another way of putting this is that the process of ad hoc data infrastructure expansion and incremental vendor procurement without a systemwide approach to court data infrastructure almost inevitably produces incompatible data linkages that ultimately serve to stymie court innovation.

In the section preceding, I focused on my own field observations in one particular county, but in the sections to follow, I zoom outwards to consider the general contours of case management systems and the decision that courts must make between internal and external vendor products.

B. Home-grown Data Products vs. Vendor Procurement

Case Management Systems are at the core of a court's internal functions, storing the day-to-day information that becomes the data used for justice administration, in litigation, and in academic and policy research. The abilities and benefits of a well-functioning CMS, therefore, go far beyond just making documentation digital. Former Minnesota Supreme Court Commissioner Richard Slowey emphasizes this distinction, defining a CMS as:

> [A] computerized system for case initiation, docketing case progress, calendaring, document generation, and report generation. Less apparent, yet equally important, is that a well-designed modern case management system will deliver that core functionality in a manner that provides meaningful ancillary benefits to the courts, such as more efficient data entry, more effective data retrieval, fewer data errors, better case management tools, enhanced bar and public access to case information, and others.[80]

---

[78] Judith Prieve, *Police Software Crashes Forced Dispatchers to Hand-Write Call Info, Slowed Responses*, THE MERCURY NEWS (Jun. 23, 2023), https://www.mercurynews.com/2023/06/16/police-software-crashes-forced-dispatchers-to-hand-write-call-info-slowed-responses/ [https://perma.cc/S756-EDAW].

[79] Albrecht, *supra* note 62, at 23.

[80] Richard Slowes, *Benefits of a modern court case management system*, THOMSON REUTERS 1 (2012), https://static.legalsolutions.thomsonreuters.com/index/pdf/benefits-of-a-modern-court-case-management-system-whitepaper.pdf [https://perma.cc/EWK9-MC4G].

The project of instituting a CMS in a court that does not have one, or that needs a significantly revamped one, is daunting. Some courts, like Hamilton County, have taken this project on internally with CMS's that are internally developed and over which they retain total technological control. The benefits of internal development include customization and potential cost savings. An internally developed product can cater specifically to the needs and uses of the particular court, with local technologists on site to make changes quickly. Internal development also has substantial cost-saving potential. In the case of the Hamilton County's Juvenile Court, even the most preliminary external estimates for their e-Filing system came in at hundreds of thousands of dollars.[81] And like the cautionary example of the Mark43 contract, those costs can only go up. That said, the disadvantage of an internally developed product is . . . internally developing a product. Marshaling the internal resources to build a CMS is not feasible for many local courts, absent a structural change in technology investment across the courts more broadly.

Rising to fill the need for CMS's and other data products, a number of private vendors have emerged into a lucrative marketplace. However, as courts have learned in previous painful lessons, not all data efficiency tools are good tools to invest in.[82] Importantly, if a private vendor is retained to build data products for a government entity, they should be held to the standards of data production and sharing at least commensurate with that for an internally developed product. That is, using a vendor product does not void the right of the public to access court records. I argue, however, that in practice the world of vendor systems often does not vindicate the public right to data.

Beyond the technological infrastructure problems, it is easy to see why a vendor CMS solution is tempting. Tyler Technologies, one of the biggest players in the CMS vendor space, advertises their capabilities as:

> Tyler's cutting-edge court case management systems streamline processes and eliminate mountains of paper handling. We have a track record of success across thousands of courts, from the smallest municipal court to the largest statewide systems.[83]

A turnkey solution with IT support from Tyler Technologies alleviating the strain on courts seems too good to be true. As per Tyler Technologies' own statistics, their CMS Odyssey is used in at least 600 counties in twenty-one different states

---

[81] Albrecht, *supra* note 62, at 9.

[82] *See generally* Jason Tashea, *A Human Rights Approach to Justice Technology Procurement*, GEO. L. TECH. REV. (2023), https://georgetownlawtechreview.org/a-human-rights-approach-to-justice-technology-procurement/ [https://perma.cc/9DFR-WPEU] (describing how courts have made use of a variety of algorithmic tools in ways that can inject bias into the judicial process).

[83] Courts & Justice, TYLER TECHNOLOGIES, https://www.tylertech.com/solutions/courts-public-safety/courts-justice [https://perma.cc/VF76-9WEL] (last visited Nov. 12, 2024).

including some of the biggest jurisdictions in the United States.[84] A recent survey of Tyler Technologies lawsuits identified fifty-one cases, some of which raise significant concerns that court-clients should be aware of and have contingency plans in place to mitigate. These include mistaken release of inmates, cases where inmates were wrongfully not released, wrongful arrests, inefficient eCourts systems delaying court functioning, data breaches, security glitches leading to accidental publication of over 260,000 confidential attorney discipline records, and canceled contracts and lawsuits due to bad coding.[85] In 2019, North Carolina signed a 100 million dollar contract with Tyler Technologies, but when the system debuted in 2023 (after a year-long delay and a fifteen-million dollar budget overage), it quickly became a nightmare due to "application issues and product defects."[86] Soon after, a class-action lawsuit was filed alleging,

> "There have been instances where Tyler Technologies has rolled out similar software in different states over the last decade where people have been wrongfully arrested, over-detained, re-arrested on charges that they thought they were cleared," said Zack Ezor, an attorney with the North Carolina law firm Tin Fulton Walker & Owen, and a representative of the two named plaintiffs in the class-action suit.[87]

The lawyers went on to call the problems with Tyler Technologies "entirely foreseeable."[88] Courts need to establish significant safeguards and responsible system roll-out plans, assuming that there will be challenges in adopting a new product, as there almost always are. It is no longer justifiable for courts to be unprepared for a known slate of potential fail points when such problems are well documented and well litigated.

C. Diversity in Case Management Systems

While Tyler Technologies is a big player in the CMS and other justice data products markets, they are not the only one. In order to partially evaluate the terrain

---

[84] *Tyler Technologies Provides Update on Odyssey Court Case Management Implementation in Shelby County*, TYLER TECHNOLOGIES (Dec. 07, 2016), https://investors.tylertech.com/news/news-details/2016/Tyler-Technologies-Provides-Update-on-Odyssey-Court-Case-Management-Implementation-in-Shelby-County-12-07-2016/default.aspx [https://perma.cc/C3DV-KCYX].

[85] List of cases against Tyler Technologies, Inc., LAW360, https://www.law360.com/search/cases?q=%22tyler+technologies%22&submit=Submit&facet=&facet_added=&facet_removed=&range_filter=&date_start=&date_end=&per_page=100&view_style=&filter=&c_sort= [https://perma.cc/43E6-D6DQ] (last visited Oct. 4, 2024); *see* Albrecht, *supra* note 62, at 24 (referring to research conducted by Ashwin Ramaswami).

[86] Michael Hewlett & Jeffrey Billman, *A $100 Million Mess*, THE ASSEMBLY (Mar. 29, 2023), https://www.theassemblync.com/politics/courts/a-100-million-mess/ [https://perma.cc/63X5-TJBN].

[87] Rusty Jacobs, *Company Behind A Digital Court Filing System in North Carolina Now Faces A Class-Action Lawsuit*, WUNC N.C. PUB. RADIO (May 24, 2023), https://www.wunc.org/law/2023-05-24/company-behind-a-digital-court-filing-system-in-north-carolina-now-faces-a-class-action-lawsuit [https://perma.cc/E95D-BZBY].

[88] *Id*.

of CMS vendors, SCALES conducted an exploratory analysis of what products different courts in the State of Georgia use for case management.[89] Two companies, Icon Software and MicroPact (formerly called Iron Data) are the architects of many of the data portals in Georgia, meaning that the interfaces are consistent within each company's client base.[90] Tyler Technologies via re:SearchGA also has significant clients across the state, listing nineteen locations as currently integrated and fifteen in progress, though for many of those courts, integration applies only to cases dating back to 2022.[91,92] Four courts had their own data products: Clayton, Cobb, Hall, and Houston, though Clayton is on Tyler Technologies' in-progress list. The ubiquity of vendor products in the CMS space is not unique to Georgia. Tyler

[89] SCALES refers to Systematic Content Analysis of Litigation EventS – OKN, a non-profit that builds open-source data products to help make the justice system more open and transparent. Home Page, SCALES-OKN, https://scales-okn.org/ [https://perma.cc/PZP3-GJ6U] (last visited Nov. 13, 2024). SCALES, and some of the specific data products referenced, are described in detail in Part IV of this Article. This author is also a member of the SCALES development team.

[90] We confirmed the use of Icon Software in thirty-nine courts superior courts, with many of the counties using the same company: Banks, Bleckley, Burke, Butts, Chattahoochee, Chattooga, Cherokee, Clarke, Columbia, Crawford, Dawson, Dodge, Fayette, Floyd, Franklin, Grady, Habersham, Harris, Hart, Jackson, Madison, Meriwether, Miller, Montgomery, Morgan, Oglethorpe, Paulding, Peach, Pickens, Pulaski, Putnam, Rabun, Richmond, Talbot, Taylor, Telfair, Toombs, Troup, and Wilkinson. We confirmed the use of MicroPact in ten courts, with many of the counties using the same company: Baldwin, Bryan, Camden, Carroll, Catoosa, Coweta, Effingham, Henry, Liberty, and Newton. Methodologically, these determinations were made by visiting court data portal websites.

[91] The following courts were labeled by Tyler Technologies as "Integrated:" Bib, Bryan, Camden (superior), Chatham, Cherokee, Coweta (superior), DeKalb, Elbert, Forsyth, Fulton, Glynn, Gwinnett, Long, Lowndes, Oconee, Rockdale, Screven, Spalding, and Washington; Labeled by Tyler Technologies as "Coming Soon," "E-filings Only," or "In Progress": Camden (state), Clayton, Coweta (state), Decatur, Douglas, Effingham, Evans, Floyd, Henry, Jenkins, Liberty, McIntosh, Pickens, Pierce, and Thomas. Tyler Technologies, *Integrated Courts*, RE:SEARCHGA, https://researchga.tylerhost.net/CourtRecordsSearch/Home#!/integratedCourts [https://perma.cc/2ZCA-J43G] (last visited Oct. 4, 2024).

[92] Notably, Tyler Technologies acquired MicroPact in 2019, but have not, as of the time of this writing, updated all of the MicroPact portals forward. *Tyler Technologies to Acquire MicroPact from Arlington Capital Partners*, TYLER TECHNLOGIES (Feb. 01, 2019), https://investors.tylertech.com/news/news-details/2019/Tyler-Technologies-to-Acquire-MicroPact-from-Arlington-Capital-Partners-02-01-2019/default.aspx [https://perma.cc/B2CG-QD3F].

Technologies product, Odyssey, is deployed statewide or widely across the state in Indiana,[93] Texas,[94] North Carolina,[95] Kansas,[96] Maine,[97] and others.

This raises a larger question about system unification. The appeal of a unified system—namely, compatibility across jurisdictions within the state both for data transfer and staff training—is obvious. However, an under-considered caveat is whether system unification is actually good for individual units. Tennessee has recently passed House Bill 2930[98] creating a unified court system and accompanying Case Management System, appropriating $75 million to execute the new system.[99] House Majority Whip Johnny Garret lauded the move, saying,

> We have more than 300 court systems in the state and they all operate on different systems. . . . Technically, they can't talk to one another. This unified system will greatly improve the communication between our clerks while also allowing us to collect data on what crimes are committed and where they are taking place.[100]

Despite the positives of a unified system in theory, I found in my research in Hamilton County that some counties across the state are not happy in practice. Among these are the largest courts in the state, who already have customized and feature-rich CMS's that they would like to keep. Communicated worries include losing features due to a standardization process meant to encompass 300 courts with a small budget and limited access to technological support or development both for small jurisdictions and from a hypothetical vendor.[101] Hamilton County and their municipal courts also reported concerns about the ability of vendor products to

---

[93] Jordan Morey, *On the Same Page: Odyssey CMS Now in All 92 Counties*, THE IND. LAW. (Feb. 2, 2022), https://www.theindianalawyer.com/articles/on-the-same-page-odyssey-cms-now-in-all-92-counties [https://perma.cc/ET5H-72GL].

[94] *Tyler Technologies Successfully Implemented Odyssey at More Than 30 Jurisdictions in the Past 18 Months*, BUS. WIRE (Jul. 7, 2015), https://www.businesswire.com/news/home/20150707005286/en/Tyler-Technologies-Successfully-Implemented-Odyssey-at-More-Than-30-Jurisdictions-in-the-Past-18-Months [perma.cc/7UJ5-SHFF].

[95] Hewlett & Billman, *supra* note 86.

[96] Sophia Fox-Sowell, *Kansas' Electronic Courts System Still Down Weeks After Cyber Incident*, STATESCOOP, (Nov. 7, 2023), https://statescoop.com/kansas-electronic-courts-system-still-down-weeks-after-cyber-incident/ [https://perma.cc/BAK5-4HE8].

[97] James McCarthy, *Tyler Technologies Lands $16.8M Contract with Maine's Court System*, MAINEBIZ (Dec. 23, 2016), https://www.mainebiz.biz/article/tyler-technologies-lands-168m-contract-with-maines-court-system [https://perma.cc/BV65-M7TM].

[98] H.R. 2930, 112th Gen. Assemb., Reg. Sess. (Tenn. 2022).

[99] *Unified Court System Streamlines Judiciary*, COFFEE CNTY BROAD. WMSR-AM OPIF (May 30, 2024), https://thunder1320.com/unified-court-system-streamlines-judiciary/ [https://perma.cc/L52A-N2H4].

[100] *Id.*

[101] Note that $75 million is allocated for the whole process, from planning to execution, not just for a vendor contract.

throttle public access by virtue of removing the ability of courts to control their own data and instead placing it in the pipeline of a commercial vendor.[102]

IV. PATHWAYS TO INFRASTRUCTURE DEVELOPMENT

In the next two sections of this Article, I offer answers to the question: what should courts do? I open with a discussion of knowledge graph architecture as a new way of conceptualizing data linkages in an open justice system, including the discussion of data products and a current use case. Next, I consider the potential harms of data linkages and the measures that need to be taken to protect individuals in an open data system.

A. Knowledge Graphs as Court Architecture

Open Knowledge Networks are a combination of knowledge graphs and open-source data linkages. Knowledge graphs provide a way to combine and enrich data from a wide array of sources using a graph-like infrastructure.[103] Put more simply, they "describe[] objects of interest and connections between them." [104] Knowledge graphs have been around in some form since the 1970s, but it is only more recently that they have entered the technological milieu in the form of the Google Knowledge Graph in 2012, where Google changed the means of searching for data from strings to entities.[105] However, a more specific definition of knowledge graphs has recently emerged, generally requiring a graph architecture and some sort of internal schema that makes entities within the graph definable and connectable.[106] Examples of knowledge graphs that are available to use now include both closed, commercialized graphs, which utilize data from private sources, and open knowledge graphs, which make use of publicly available data. Examples of

---

[102] In my observations at Hamilton County, I saw this firsthand. Where the criminal division uses CJUS, the civil division instead uses a vendor product accessible to the public via Tennessee CaseFinder. The interested public is not the only user of Tennessee CaseFinder, which is also used extensively by Legal Aid. When the vendor noticed an unusually high use rate of the public portal, they responded by throttling the portal, creating barriers to access for the general public and Legal Aid. While the portal was eventually restored, this situation exemplifies the fragility of public access.

[103] Aidan Hogan, *Knowledge Graphs: A Guided Tour*, *in* 2022 INTERNATIONAL RESEARCH SCHOOL IN ARTIFICIAL INTELLIGENCE IN BERGEN 1, 1 (Camille Bourgaux et al. eds. 2022).

[104] Natasha Noy et al., *Industry-Scale Knowledge Graphs: Lessons and Challenges: Five Diverse Technology Companies Show How It's Done*, 17 QUEUE 48, 48 (2019).

[105] Amit Singhal, *Introducing the Knowledge Graph: Things, Not Strings*, GOOGLE BLOG, (May 16, 2012), https://blog.google/products/search/introducing-knowledge-graph-things-not/ [https://perma.cc/8E2N-DJ6Z] (The movement from the 'old internet' means of searching where the functionality is to match a string of text exactly or within a margin of error to another string of text. This method does not account for the meaning of that string, just the characters that compose it. In contrast, an entity-focused search seeks to define the entity, determine what it means, and then provide the results that have that meaning or a closely related meaning.)

[106] Hogan, *supra* note 103, at 2.

companies with closed knowledge graphs include Airbnb, Amazon, and eBay.[107] While knowledge graphs have come quite a long way, they are not without challenges, which include the scalability of the graphs, the process of obtaining unstructured information to populate them, and future-thinking concerns about how to update them with new and changing knowledge.[108] This issue resonates particularly strongly with the court data example, where if information is de jure public, but not de facto accessible,[109] knowledge graphs cannot be populated effectively.

The 'open' part of open knowledge networks is more than just a general imperative for open and accessible data, rather it is also a specific initiative between the National Science Foundation and the U.S. Government. In 2017, the National Science Foundation acknowledged the need for and utility of an Open Knowledge Network.[110] Two years later, they funded twenty-one Open Knowledge Network projects via the Convergence Accelerator Track A.[111] Following the completion of these multi-million-dollar interdisciplinary projects, the National Science Foundation convened a design sprint to develop a roadmap for the future of Open Knowledge Networks with the Office of Science and Technology Policy, culminating in a 204-page report and a wave of funding for a new set of Open Knowledge Network projects.[112]

An important feature of an Open Knowledge Network that spans many types of data, types of industry, and social problems is its ability to expand and harmonize across itself. The unifying method of doing this is via the National Information Exchange Model ("NIEM"). Formally launched in 2005, NIEM "defines agreed-

---

[107] Spencer Chang, *Scaling Knowledge Access and Retrieval at Airbnb*. MEDIUM: THE AIRBNB TECH BLOG, (Sept. 4, 2018), https://medium.com/airbnb-engineering/scaling-knowledgeaccess-and-retrieval-at-airbnb-665b6ba21e95 [https://perma.cc/VL3E-3DHA]; Arun Krishnan, *Making Search Easier*, ABOUT AMAZON (Aug. 17, 2018), https://www.aboutamazon.com/news/innovation-at-amazon/making-search-easier [https://perma.cc/9X2W-UCHD]; R. J. Pittman et al., *Cracking the Code on Conversational Commerce*, EBAY: STORIES, (Apr. 6, 2017), https://www.ebayinc.com/stories/news/cracking-the-code-on-conversational-commerce/ [https://perma.cc/35DG-J5MA].

[108] Noy et al., *supra* note 104.

[109] Albrecht & Filip, *supra* note 9, at 6-7.

[110] *Harnessing the Big Data Revolution*, U.S. NAT'L SCI. FOUND., https://www.nsf.gov/news/special_reports/big_ideas/harnessing.jsp [https://perma.cc/28A9-2R52] (last visited Nov. 12, 2024); Chaitan Baru et al., *Open Knowledge Network Roadmap: Powering the Next Data Revolution*, U.S. NAT'L SCI. FOUND. CONVERGENCE ACCELERATOR 13 (Sept. 2022), https://nsf-gov-resources.nsf.gov/2022-09/OKN%20Roadmap%20-%20Report_v03.pdf [https://perma.cc/HG9G-J4FY].

[111] *NSF's Convergence Accelerator Track A: Open Knowledge Network (Harnessing the Data Revolution) Phase 1 Awardees*, U.S. NAT'L SCI. FOUND., https://nsf-gov-resources.nsf.gov/2023-01/Track_A-OKN%20Harnessing_the_Data_Revolution_Phase%201%20Awardee%20List.508.pdf?VersionId=YdsrHWwRCRU_wnq_cfeXFJZlNyEAXVov [https://perma.cc/RLA7-F84Q] (last visited Aug. 01, 2024) [hereinafter *NSF's Convergence Accelerator Track A*]; Baru et al., *supra* note 110, at 13.

[112] *NSF's Convergence Accelerator Track A*, *supra* note 111; About the Office of Science and Technology Policy, THE WHITE HOUSE (Aug. 01, 2024), https://www.whitehouse.gov/ostp/ [https://perma.cc/RT6T-QFRB]; Baru et al., *supra* note 110, at 13.

upon terms, definitions, relationships and formats—independent of how information is stored in individual systems—for data being exchanged."[113] NIEM places language into two categories: core elements that are used across all of NIEM, and community-specific elements that are specific to individual parts of the graph that can be aligned to the rest of the ontology.[114] For example, if two different groups were trying to name a person/place/concept, they might just naturally use slightly different labels to mean the same thing. In NIEM, that problem is avoided by using the NIEM naming specification.[115] Of course, this requires that data holders desire and ultimately do the work to make their data holding aligned with NIEM.

As creators, producers, and maintainers of public data, courts should aim to realign their data infrastructure to the Open Knowledge Network. This seems far afield from the current terrain of court data and court data infrastructure, but there are multiple use cases that currently exist, suggesting that such a future is possible. Through two case studies, I will first demonstrate how court data can be situated with the Open Knowledge Network and then provide an example of a court transitioning its internal systems with knowledge graph infrastructure.

1. Systematic Content Analysis of Litigation EventS (SCALES)

Originating from one of the National Science Foundation's Convergence Accelerator projects, the Systematic Content Analysis of Litigation EventS ("SCALES") is a nonprofit dedicated to building open knowledge graph data explorers designed to make the judicial system more transparent via open, accessible, and linked court records.[116] The goals of SCALES are explicitly related to the Open Knowledge Network, describing the requirements of an open knowledge network as something "freely available to all stakeholders, including the researchers who will help push this technology further. It is a nonproprietary public-private development effort that spans the entire data science community. The result of an OKN is an open, shared infrastructure."[117]

---

[113] *NIEM Model*, NIEMOPEN, https://www.niem.gov/about-niem/niem-model [https://perma.cc/AP33-PBFW] (last visited Nov. 12, 2024).

[114] *NIEM Model*, *supra* note 113.

[115] Naming and Design Rules Directory, NIEM, https://reference.niem.gov/niem/specification/naming-and-design-rules/ [https://perma.cc/QUW5-CV4Y] (last visited Nov. 12, 2024).

[116] *See generally*, Adam R. Pah et al., *How to Build a More Open Justice System*, 369 SCIENCE 134, 136 (2020) (describing the need for a more open and transparent judicial system, speaking in particular about their first use case with federal court records).

[117] Schwartz et al., *supra* note 1, at 37-38.

SCALES started with the Federal Courts and the PACER system as a first use case but has since begun an expansion into state courts and criminal legal system data.[118] SCALES describes their product and process as follows:

> SCALES uses AI tools to create a platform that enables systematic analysis of court records.[119] This platform is made publicly available via a data explorer.[120] Crucially, users can take full advantage of the data explorer without any computer programming knowledge. The data explorer is designed to accept common-language queries and questions. The underlying data powering the data explorer is drawn from PACER using software that automatically downloads queries, dockets, case summaries, and documents.[121] Importantly, the goal of SCALES is not to serve as a financial intermediary to PACER. Instead, SCALES extracts, transforms, and enriches PACER data to make it amenable to nuanced analysis and accessible to everyone.[122]

Figure 1 (below) depicts the underlying knowledge graph structure for the SCALES OKN Data Explorer (the "Data Explorer"). The graph architecture is specifically designed to be expansive at the upper levels, allowing for integration of additional use cases.[123] In the secondary tier of the graph, there is substantial infrastructure dedicated to alignment and concept elucidation specifically compatible with the NIEM Model.[124] Data sources enter the pipeline after conceptual domain ontology mapping[125] and are then routed into Application

---

[118] *See Award Abstract # 2033604, A1: Systematic Content Analysis of Litigation Events (SCALES) Open Knowledge Network to Enable Transparency and Access to Court Records*, U.S. NAT'L SCI. FOUND., https://www.nsf.gov/awardsearch/showAward?AWD_ID=2033604&HistoricalAwards=false [https://perma.cc/3DML-C6W3] (last visited Nov. 12, 2024); *see also Award Abstract # 2333803, Proto-OKN Theme 1: An integrated platform to connect criminal justice data across data silos,* U.S. NAT'L SCI. FOUND., https://www.nsf.gov/awardsearch/showAward?AWD_ID=2333803&HistoricalAwards=false [https://perma.cc/4ARP-XTRK] (last visited Nov. 12, 2024).

[119] For a detailed description, see *SCALES OKN Documentation*, SCALES, https://docs.scales-okn.org/ [https://perma.cc/2WJU-D76W] (last visited June 24, 2024).

[120] To access the data explorer, visit *Transforming the Accessibility and Transparency of Federal Courts*, SCALES, https://scales-okn.org [https://perma.cc/7JV3-JFR7] (last visited April 2, 2024).

[121] All SCALES software is produced under a GPL license and is available for use. *See generally* SCALES, GITHUB, https://github.com/scales-okn [https://perma.cc/6842-5D6L] (last visited April 2, 2024) (hosting the SCALES software along with a full suite of documentation).

[122] Schwartz et al., *supra* note 1, at 38-39.

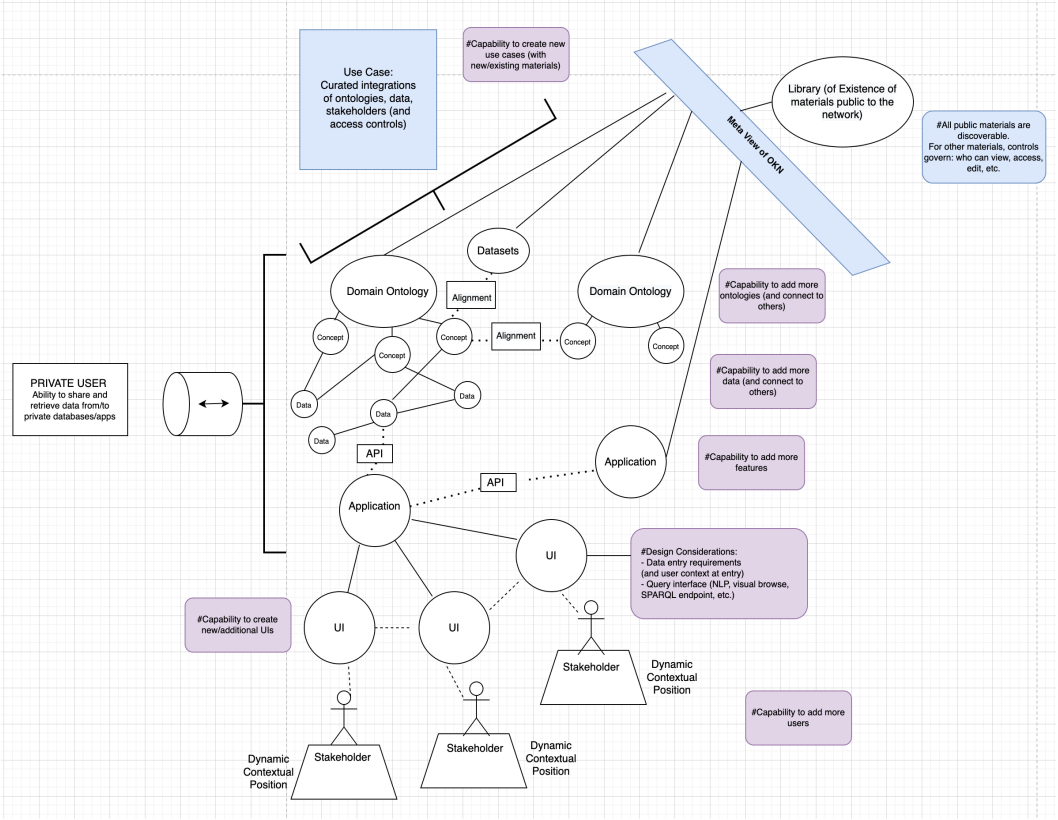
[123] *SCALES OKN Documentation*, *supra* note 119. The upper levels make up the generalized central section of the graph, responsible for housing, connecting, and manipulating data, while the secondary tier toward the outside of the graph is more specific, responsible for facilitating applications of the processed data.

[124] The Open Knowledge Network program specifically funds a NIEM ontology team and the Fabric Integrating Networked Knowledge ("FRINK") to ensure that additionally developed Knowledge Graphs are compatible with existing ontologies.

[125] Domain ontology mapping is the process of establishing relationships between specific data and generalized concepts and terms. The resulting 'map' is useful for data distribution and analysis. *See Ontology Mapping*, SCIENCE DIRECT, https://www.sciencedirect.com/topics/computer-science/ontology-mapping [https://perma.cc/WF42-QVBU] (last visited Nov. 13, 2024).

Programming Interfaces or other methods of delivery into an application.[126] The next layer of the knowledge graph more directly engages end users. Building a complex and involved data infrastructure might be scientifically and analytically sound, but it does not successfully execute the directive of 'open and available to everyone' if regular people cannot figure out how to use it. Therefore, there is considerable effort and energy invested in the development of a user interface, whatever the end product is, that someone interested in the data would actually want to use.[127]

Figure 1: SCALES OKN Data Explorer Knowledge Graph Structure



For SCALES' first project with federal court records coming from PACER, you can trace the graph structure all the way to the end product. Taking federal court records as a use case, SCALES first had to develop the local domain ontology, determining what concepts and types of data information would be retained and

---

[126] *See generally* Michael Goodwin, *What is an API (application programming interface)?*, IBM (Apr. 9, 2024), https://www.ibm.com/topics/api [https://perma.cc/P9MW-WDAX] (defining Application Programming Interfaces, APIs, as intermediaries that help different applications talk to each other, allowing applications to exchange data and features according to a set of rules or protocols)

[127] The usability of the final data product is of utmost importance. Many of the examples here, like PACER and the internal CJUS system in Hamilton County, illustrate this necessity of considering user experience in designing court data products. Not only do these products need to be well designed for their initial implementation, but they also need to be regularly updated as best practices and technological capacities are also updated.

enriched from the PACER data.[128] This enrichment process consumed significant time and involved the knowledge and efforts of many domain experts across law and the social sciences to try and build an end product that was the most useful.[129] The end result was the Data Explorer, a free, shareable, no-code-required data tool where users can ask questions about federal court records and receive visualizations of the answers and cases that contributed to those visualizations.[130]

This example and the new project SCALES has undertaken to build data explorers for state and county criminal courts demonstrate how the Open Knowledge Network can take data out of existing court systems and create enriched, open-access analytic tools.[131] Pivotal to SCALES' success in this work is the ability to get data from the courts. After this data is obtained, data enrichment will be the most time-consuming part of the SCALES process. If courts observed more of the best practice recommendations in data infrastructure building,[132] the time necessary to develop these dynamic tools would be significantly shortened.

2. Sonoma County, California

The SCALES example is one where data comes out of courts to enter the Open Knowledge Network. In contrast, Sonoma County, California provides an example of a county redesigning its internal architecture to be knowledge network compatible by designing an Integrated Justice System ("IJS") to harmonize formerly incompatible Case Management Systems.

Sonoma describes IJS, stating, "[T]he data and interoperability backbone that connects the Sheriff, District Attorney, Public Defender, Courts and Probation in a County. These agencies typically run different Case Management Systems and need the IJS for data exchange and workflow orchestration."[133] The benefits of IJS as enumerated by Sonoma County include a better user experience to find needed data, modernization without disrupting essential court functions, centralizing technology to lessen individual technology costs, increasing cross-agency collaborations, and providing opportunities for phased transition to the new knowledge network.[134]

Importantly, Sonoma's initiative also reaches outward to be NIEM-compatible, including requiring NIEM alignment for all new vendor products as per contractual

---

[128] *See* Schwartz et al., *supra* note 1, 41-43.

[129] See *Our Team*, SCALES, https://scales-okn.org/team-2/ [https://perma.cc/7EM2-GPNG] (last visited Nov. 3, 2024), for a list of experts involved in the SCALES' project.

[130] To access the data explorer, visit *Sign Up*, SCALES, https://satyrn.scales-okn.org/ [https://perma.cc/64BE-6K79] (last visited Nov. 5, 2024).

[131] *Award Abstract # 2333803*, *supra* note 118.

[132] *See* discussion *infra* Part V.

[133] Presentation to the U.S. National Science Foundation, Sonoma IJS Assessment Summary 3 (July 11, 2024) (on file with the Columbia Science and Technology Law Review).

[134] *Id.*

language.[135] Sonoma's success demonstrates that open knowledge graph structure is possible at the county level and that we need not exclude county courts from discussions of data innovation.

B. Mitigating Harm while Increasing Transparency

A paradigm shift to an Open Knowledge Network, or even a partial realignment to a more open justice system, necessarily prompts careful consideration of the potential benefits and harms of making data more open. While a majority of court records are public, the fact that most of the public cannot meaningfully access them functionally alters their status to non-public records.[136] In this reality, we have to carefully consider how making court records actually public would affect the people described in them. Court records contain a lot of personal information, from names and addresses to criminal actions, even possibly including case-related medical information. Court data can therefore be very sensitive and cause reputational harm or be used to discriminate against individuals in society in a number of ways.[137] However, as this article has argued and analyzed, court data opacity leaves us with a closed justice system that is also ripe for sheltering biases, discrimination, and distrust.

It is also worth noting that just because the general public cannot meaningfully access court records does not mean they are completely inaccessible. Some actors or industries do reliably use court records and their derivatives. Consider background checks, a $2 billion industry that is part of most hiring practices.[138] This practice hovers somewhere between protective and punitive. Federal and state laws require background checks in some industries, like those concerning work with children.[139] They also require notification and proper procedures when

---

[135] Email from Sarkar Suhankar, Managing Partner, Ernst & Young LLP, to Kat Albrecht, Assistant Professor, Georgia St. U., Andrew Young Sch. of Pol'y Stud. (July 11, 2024, 03:22 PM) (on file with the Columbia Science and Technology Law Review).

[136] Albrecht & Filip, *supra* note 9, at 6-7.

[137] Lynn M. LoPucki, *Court-System Transparency*, 94 IOWA L. REV. 481, 516-17 (2009); Kaitlyn Filip & Kat Albrecht, *Regulating Harm: Tensions Between Data Privacy and Data Transparency*, J. REG. COMPLIANCE 115, 120-21 (2022).

[138] Marina Duane et al., *Criminal Background Checks*, URB. INST. JUST. POL'Y CTR. 1, 11 (Nov. 2017), https://www.urban.org/sites/default/files/publication/88621/2001174_criminal_background_checks_impact_on_employment_and_recidivism_2.pdf [https://perma.cc/FH2F-ZH92].

[139] *See* 34 U.S.C. § 20351(a), requiring background checks for federal employment:

> (a) In general
>
> (1) Each agency of the Federal Government, and every facility operated by the Federal Government (or operated under contract with the Federal Government), that hires (or contracts for hire) individuals involved with the provision to children under the age of 18 of child care services shall assure that all existing and newly-hired employees undergo a criminal history background check. All existing staff shall receive such checks not later than May 29, 1991. Except as provided in subsection (b)(3), no additional staff shall be hired without a check having been completed.

intending to run a background check.[140] However, requiring background checks for employment in general also serves as a significant sociological barrier for the 33% of Americans with criminal records attempting to obtain employment, even when such employment is required by the court.[141]

Commercial entities have also weaponized public records against individuals for profit, further demonstrating the reality that public records are not truly public in their current form. Otherwise, such commercial schemes would not provide a novel product. Take the mugshot publishing industry as an example. With over 4.5 million mugshots released per year as part of public court records, dozens of websites came to prominence in the mid-2010s to host and disseminate these images.[142] These websites, in addition to profits via advertising, would charge the individuals a fee to remove their photographs from the site in what the American Bar Association called, "an online extortion scheme."[143] Attempts to stifle the mugshot publishing industry were challenging. One Georgia District Attorney noted that arbitrarily charging for mugshot removal doesn't fit the legal definition of extortion because the photos are public record; "it's wrong but not a violation of the criminal laws."[144] Since then, at least thirteen states have introduced bills attempting to curtail the practice, and several high-profile lawsuits have been filed, including a notorious class action lawsuit against the owners of Mugshots.com.[145]

See also VA. CODE ANN. §§ 21.1-289.034 to -289.041 (2023); FLA. STAT. § 456.0135 (2023); WASH. REV. CODE § 43.215.270 (2023); OHIO REV. CODE ANN. § 3319.391 (LexisNexis 2023).

[140] *See* EQUAL EMPLOYMENT OPPORTUNITY COMMISSION, EEOC-NVTA-0000-38, BACKGROUND CHECKS: WHAT EMPLOYERS NEED TO KNOW (Mar. 11, 2014), https://www.eeoc.gov/laws/guidance/background-checks-what-employers-need-know [https://perma.cc/5NVG-ZL4D] (listing federal laws and regulation relevant to background checks in employment).

[141] Duane, *supra* note 138, at 2 n. 14 ("In 2014, there were 105,569,200 individuals in state criminal history files. The US Census Bureau estimates that in July 2014, the total United States population was 318,857,056. This equates to approximately 33 percent of Americans with a criminal record (BJS, Survey of State Criminal History Information Systems, 2014)").

[142] Sarah E. Lageson et al., *Digitizing and Disclosing Personal Data: The Proliferation of State Criminal Records on the Internet*, 46 LAW & SOC. INQUIRY 635, 637 (2021); Eumi K. Lee, *Monetizing Shame: Mugshots, Privacy, and the Right to Access*, 70 RUTGERS U. L. REV. 557, 566 (2018).

[143] Jason Tashea, *2018's most important legal tech stories*, ABA J. (Dec. 21, 2018, 6:30 AM), https://www.abajournal.com/news/article/toptentech [https://perma.cc/7N74-JUXW].

[144] Josh Green, *Mugshots Inc: 'Legalized extortion' or Constitutional privilege?*, GWINNETT DAILY POST (July 22, 2012), https://www.gwinnettdailypost.com/archive/mugshots-inc-legalized-extortion-or-constitutional-privilege/article_e5655e0d-0a1e-5b28-b1e1-364383c0cfa4.html [https://perma.cc/R8ZG-V7WD].

[145] *See* Notice of Removal, Ex. 1 (Complaint), *Gabiola v. Keesee*, No. 1:16-cv-02076 (N.D. Ill. filed Feb. 9, 2016), ECF. No. 1. The states included in this list are Arizona (H.B. 2191, 44th Leg., Reg. Sess. (Ariz. 2019)), California (S.B. 1027, 2013-14 Leg., Reg. Sess. (Cal. 2014)), , Colorado (H.B. 14-1407, 69th Gen. Assemb., 2d Reg. Sess. (Colo. 2014)), Connecticut (Comm'r of Pub. Safety v. Freedom of Info. Comm'n, 312 Conn. 513 (2014) (holding that law enforcement agencies are required to disclose a person's "record of arrest," which need not include mugshots)), Florida (S.B. 118, 2017 Leg., Reg. Sess. (Fla. 2017)), Georgia (H.B. 150, 152nd Gen.

This leaves courts in a tricky situation, where we know that public records can be quite damaging even though they retain a status as constitutionally public. Attempting to resolve this tension, we can further subdivide public data into two categories: data that is germane to the public interest and data that is not. When designing the SCALES-OKN, the SCALES team applied this concept and therefore decided to leave the names of lawyers and judges in the data but replaced the names and addresses of individual litigants with a hashed number.[146] Their rationale was that lawyers and judges are acting in their official capacities in the public arena, whereas individual litigants are not. Now, because PACER records are technically publicly accessible, a member of the public could reidentify a record by looking up a case and putting the name back into it. However, SCALES' choice to provide some type of blinding makes it substantially more difficult for a commercial-type entity to add back individually identifying information en masse. These types of questions about data privacy versus data transparency will become even more salient as data becomes increasingly digitized, regardless of whether courts choose to redesign their own infrastructure or not. Taking intentional control over infrastructure management and implementing best practices will help alleviate some of the potential harms of open data before the consequences are fully felt by members of the general public.

V. BEST PRACTICES RECOMMENDATIONS

To that end, I present here a series of best practices recommendations, with the intention of offering possibilities to courts with varying levels of resources and at different places in the data infrastructure development process. These recommendations are derived from the research cited here, from my own field work, and from lessons learned during the development of the SCALES Open Knowledge Network.

In an ideal world, courts should make a large, intentional investment in data infrastructure. Similar to the work conducted in Sonoma County, California, this investment should extend outside the boundaries of the Case Management System in order to create cohesive connections between data products in ways that are amenable to expanding shared infrastructure.[147] In this higher-investment system, the infrastructure of a knowledge graph to facilitate the responsible delivery of legal system information to the public is a feature, not a difficult accommodation.

Assemb., Reg. Sess. (Ga. 2013), Missouri (H.B. 1665, 97th Gen. Assemb., 2d Reg Sess. (Mo. 2014), New Jersey (A.B. 2906, 2012-13 Leg., Reg. Sess. (NJ 2013), Oregon (H.B. 3467, 77th Leg. Assemb., Reg. Sess. (Or. 2013), South Carolina (S. 255, 2015-16 Leg., 121st Sess. (S.C. 2016), Texas (S.B. 1289, 83rd Leg. (Tex. 2013), Utah (H.B. 408, 60th Leg., Gen. Sess. (Utah 2013), and Virginia (S.B. 720, 2015 Leg., Reg. Sess. (Va. 2015)).

[146] For a detailed description regarding redaction, see Scott Daniel, *PACER redaction script*, SCALES (Jan. 9, 2024), https://docs.scales-okn.org/redaction/ [https://perma.cc/G8XT-7VYX]. For more information about SCALES, including user guides and additional documentation, see *SCALES OKN Documentation*, SCALES, https://docs.scales-okn.org/ [https://perma.cc/2WJU-D76W].

[147] Presentation to the U.S. National Science Foundation, *supra* note 133, at 3.

However, if a court wanted to participate in the open knowledge network space without completely abandoning the systems and structure they already have, realigning their internal system definitions and concepts to the NIEM ontology is another method of alleviating some of the data strain when it comes to providing data for case proceedings, evidence-based decision-making, and public access.

Both of these higher-investment suggestions privilege cross-unit communication to develop shared infrastructure. This requires courts to think beyond existing silos of data responsibility and work across units and organizations with the goal of intentional system design. In most courts, there is not currently a single individual for whom that is recognizably 'their job.' In a higher-investment system, it does become someone's job. That is, an investment in data infrastructure as court architecture is not just investment in technology products, it is also an investment in *people*. In Hamilton County, Tennessee, the greatest strength of their data innovation work is the diversity of local leaders who are invested in data improvement.[148] This coalition crosses partisan lines, organizational lines, and units to form a group of leaders who agree that a more efficient data system is better for all of them.

Much of this Article focused on Case Management Systems as the backbone to court data infrastructure because not all courts have established strong connections between the police or corrections and court data systems. If these system linkages are too resource-intensive, there are still things that a court can do today to improve their data infrastructure even within the confines of the Case Management System. Developing strong, well-documented, and institutionally known policies about data access and new data product procurement is necessary for all courts. Documentation, and sharing that documentation, is the first step to building institutional knowledge around the data system and exactly how it functions. Policies and procedures around who can access data and how errors in the data are resolved are absolutely necessary as data becomes the terrain of litigation and policy decision-making. Developing these policies and procedures also facilitates internal justice administration, helping ensure that internal records are as accurate as possible and keeping a chain of custody intact for any confusion that may arise.

Finally, even courts that lack the ability to revise and realign current systems can take steps to ensure that future data product procurement embodies best practices around data infrastructure and data connectivity. Courts should ensure that vendor procurement conversations involve representatives of departments and units that will be affected by the product change, including those outside the specific unit or division acquiring the new data product. This means consulting with technologists and having enough knowledge of the internal data process to know where the data comes from and where the data needs to go next before implementing the new vendor product. Increased communication across similarly situated courts could alleviate some of this strain, by which courts can share their own pain points and best practices. Vendor acquisitions should be thoroughly vetted, including as to how the new technology may affect access to historical data.

---

[148] Albrecht, *supra* note 62, at 6, 8-10.

Proactive consideration of consequences will help ensure that the future of data infrastructure in courts is better than it is today.

VI. CONCLUSION

As courts find themselves in the era of digital storage and data law, they have an opportunity to make intentional choices about modernizing, realigning, and recreating internal infrastructure around court data. Even if court data is not particularly useful, the general public has a constitutional right to access court records. Courts have recognized this right via legal decisions but have not sufficiently vindicated it due to systematic barriers preventing data access—even in situations where tools exist that theoretically make accessing court records easier. Even so, court data has become vastly more important in litigation and even internally to courts as data infrastructure becomes part of justice administration. Innovative data infrastructure options like open knowledge graphs could become a viable solution for courts that need to provide meaningful public access while also improving and streamlining their internal systems. Even if courts cannot commit to a full infrastructure redesign, there are still necessary steps courts need to take by creating policies and processes around data, documenting the structure of their internal systems, and requiring newly acquired vendor products to adhere to higher standards and best practices.