

Georgia State University

## ScholarWorks @ Georgia State University

---

CJC Publications

Department of Criminal Justice and  
Criminology

---

7-2-2022

### Public Records Aren't Public: Systemic Barriers to Measuring Court Functioning & Equity

Kat Albrecht

*Georgia State University*, [kalbrecht@gsu.edu](mailto:kalbrecht@gsu.edu)

Kaitlyn Filip

*Northwestern University*

Follow this and additional works at: [https://scholarworks.gsu.edu/cj\\_facpub](https://scholarworks.gsu.edu/cj_facpub)



Part of the [Communication Commons](#), and the [Criminology and Criminal Justice Commons](#)

---

#### Recommended Citation

Albrecht, Kat and Filip, Kaitlyn, "Public Records Aren't Public: Systemic Barriers to Measuring Court Functioning & Equity" (2022). *CJC Publications*. 41.

[https://scholarworks.gsu.edu/cj\\_facpub/41](https://scholarworks.gsu.edu/cj_facpub/41)

This Article is brought to you for free and open access by the Department of Criminal Justice and Criminology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in CJC Publications by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact [scholarworks@gsu.edu](mailto:scholarworks@gsu.edu).

# **PUBLIC RECORDS AREN'T PUBLIC: SYSTEMIC BARRIERS TO MEASURING COURT FUNCTIONING & EQUITY**

**KAT ALBRECHT & KAITLYN FILIP\***

*In a new era of computational legal scholarship, computational tools exist with the capacity to quickly and efficiently reveal hidden inequalities in the justice system. Technically, the laws exist that legally entitle the public to the requisite court records. However, the opaque bureaucracy of the courts prevents us from connecting the public to documents they technically own. We exemplify this legal ethical problem by investigating areas of law where codified protections against inequalities exist and where computational tools could help us understand if those protections are being enforced. In general, the computational requirements of such projects needn't be complex, making them even more attractive as solutions for auditing justice system processes. Using the backdrop of a national audit of public records policies to retrieve criminal jury trial transcripts, we establish the impossibility of securing the public records needed to quantify the illegal use of racially motivated peremptory strikes. We argue that the lack of opacity or availability of these policies serve as a bottleneck to the relatively simple computational process of quantifying previously unknown language and events in criminal jury trials. This article considers the ethical implications of the lack of access to records that are legally public and considers how this lack of access to records becomes an access to justice problem.*

\* Kat Albrecht is Assistant Professor at Georgia State University in the Andrew Young School of Policy Studies. Kaitlyn Filip is a Law & Humanities Fellow at Northwestern University Pritzker School of Law and a JD-PhD Student in Communication Studies: Rhetoric and Public Culture at Northwestern University. For advice and comments the authors are grateful to have presented this work at the 2022 Computational Legal Studies Conference at Singapore Management University.

## INTRODUCTION

### I. DATA RIGHTS VS. DATA ACCESS

- A. Legal Rights to Public Data
- B. Realities of Public Data Access

### II. COURT TRANSCRIPTS AS PUBLIC DATA

- A. Case Study: Cook County, Illinois
- B. Current Limitations to Knowledge due to Transcript Inaccessibility
  - 1. Court Functioning
  - 2. Access and Equity Issues

### III. AUDIT OF U.S. TRANSCRIPT AVAILABILITY

- A. Methodology
- B. Results
  - 1. Missingness
  - 2. Procedural Variation in a Legally Identical Process
  - 3. Financial Burdens and Outdated Processes
  - 4. Expert Knowledge Requirements
- C. Discussion
  - 1. Limitations

### IV. CONCLUSION

## INTRODUCTION

The nascent field of computational law has grown dramatically in recent decades, with an interdisciplinary universe of scholars tackling a variety of theoretical and empirical projects that consider both law-as-code and law-as-data. Frankenreiter and Livermore name and distinguish these two trends in computational legal analysis as the project of modeling law as a set of rules (law-as-code) versus the project of extracting information from legal text to apply to other research problems (law-as-data).<sup>1</sup> Here we focus more specifically on applications of law-as-data, but more generally argue that there is a system-level problem constraining both types of legal analysis that has yet to be dealt with.

Previously, large-scale computational legal analytics faced significant limitations in computational efficiency, cost of computing resources, and data availability, but the current

---

<sup>1</sup>Jens Frankenreiter & Michael A. Livermore, *Computational Methods in Legal Analysis*, 16 ANNUAL REVIEW OF LAW AND SOCIAL SCIENCE 39, 41 (2020).

climate of technological, scientific, and methodological innovations have made it uniquely viable to study law computationally.<sup>2</sup> Importantly, this digital era has brought with it massive increases in digital data storage and the increased attentions of social and legal scholars who endeavor to specifically harness that data using innovative computational techniques.<sup>3</sup> Consequently, a substantial amount of computational legal analysis has been undertaken in a short time. This work includes projects to obtain mass-scale legal source data, create new crosswalks of large institutional legal data, and analyze the substance of that legal data.<sup>4</sup>

At the same time as this rise of computational legal studies, various courts have publicly announced their intention to make their data more public and transparent, particularly as it

---

<sup>2</sup> See generally, RYAN WHALEN, COMPUTATIONAL LEGAL STUDIES (2020) (who presents a volume exploring and introducing computational legal studies in part due to the specific contemporaneous inflection point between law and computation today).

<sup>3</sup> Martin Hilbert & Priscila López, *The World's Technological Capacity to Store, Communicate, and Compute Information*, 332 SCIENCE 60, 64 (2011); MATTHEW J. SALGANIK, BIT BY BIT: SOCIAL RESEARCH IN THE DIGITAL AGE 3 (2019).

<sup>4</sup> See e.g. Pah, Adam R., David L. Schwartz, Sarath Sanga, Zachary D. Clopton, Peter DiCola, Rachel Davis Mersey, Charlotte S. Alexander, Kristian J. Hammond, & Luis A. Nunes Amaral. 369 *How to Build a More Open Justice System*, SCIENCE 134 (2020) (where they use open-source data from the Systematic Court And Litigation EventS data to demonstrate inconsistency in fee waiver decisions and argue for increased court data transparency); Crystal S. Yang, *Have Interjudge Sentencing Disparities Increased in an Advisory Guidelines Regime-Evidence from Booker*, 89 N.Y.U. L. REV. 1268 (2014) (who creates a cross walked dataset covering nearly 400,000 criminal defendants to study sentencing disparity); Maria-Veronica Ciocanel, Chad M. Topaz, Rebecca Santorella, Shilad Sen, Christian Michael Smith, & Adam Hufstetler, *Justfair: Judicial System Transparency through Federal Archive Inferred Records*, 15 PLOS ONE 1 (2020) (which introduces JUSTFAIR or the Judicial System Transparency through Federal Archive Inferred Records that is a large scale, cross walked, free public database of 600,000 records); Ryan C. Black & James F. Spriggs, *An Empirical Analysis of the Length of US Supreme Court Opinion*, 45 HOUS. L. REV. 621 (2008) (who use R to analyze the length of majority opinions from 1791 to 2005); Wolfgang Alschner, *The Computational Analysis of International Law*. In: RESEARCH METHODS IN INTERNATIONAL LAW 203 (2021) (who reviews computational techniques optimal for study of international law, calling for adding computer science methodology to legal scholarship); Benjamin E. Lauderdale & Tom S. Clark, *The Supreme Court's Many Median Justices*, 106 AMERICAN POLITICAL SCIENCE REV. 847 (2012) (who use kernel-weight optimal classification to weight justices' votes by both chronological closeness and substantive similarity); Daniel Martin Katz & Michael James Bommarito, *Measuring the Complexity of the Law: The United States Code*, 22 ARTIFICIAL INTELLIGENCE AND L. 337 (2014) (who offer a new framework for measuring legal complexity); Michael Evans, Wayne McIntosh, Jimmy Lin & Cynthia Cates, *Recounting the Courts? Applying Automated Content Analysis to Enhance Empirical Legal Research*, 4 J. OF EMPIRICAL LEGAL STUDIES 1007 (2007) (who test various text classification models for determining content in Supreme Court advocacy briefs); Law, David S. *The Global Language of Human Rights: A Computational Linguistic Analysis* 12 THE L. & ETHICS OF HUMAN RIGHTS 111 (2018) (who use topic modelling and natural language processing to study language similarity in national constitutions and human rights instruments); Margoni, Thomas, *Text and Data Mining in INTELLECTUAL PROPERTY LAW: TOWARDS AN AUTONOMOUS CLASSIFICATION OF COMPUTATIONAL LEGAL METHODS* (2020) (who discusses the classification of new computational legal methods and advocates for their teaching in law school curriculums).

concerns criminal courts and felony case processing. In 2018, Florida announced a new public data portal that would track criminal defendants through the system with the intention of identifying inequities.<sup>5</sup> In 2021, the state of Ohio has announced the Ohio Sentencing Data Platform in order to collect and share data about sentencing for felony cases.<sup>6</sup> Cook County, home to Chicago, Illinois and one of the largest felony criminal courts in the country, has announced a commitment to data transparency and released a substantial amount of criminal data via the Cook County State's Attorney Open Data Portal (CCODP).<sup>7</sup> Through this portal any member of the public can download and access datasets on felony case initiation, intake, diversion, disposition, and sentencing for all cases processed in Cook County Courts. Such initiatives paint a rosy picture, a picture where public data access seems to have opened up at the same time as technology has coalesced to make meaning of that data on a larger scale.

However, we argue that these initiatives have not been sufficient principally because they do not attempt to reimagine the contours of what public data actually is. These data portals have limited types of data available and leave significant gaps. Critical sources of dynamic data that are already legally public data continue to be de facto inaccessible to the public. Consequently, we argue that a critical bottleneck in both public access to data and computational legal analysis continues to be practical barriers to access that the courts have not alleviated surrounding particular types of public data.

---

<sup>5</sup> Josh Salman, *A Treasure Trove of Information, Data Collection Bill will Make Florida the Most Transparent State in the Nation*, HERALD-TRIBUNE, Mar. 12, 2018, <https://www.heraldtribune.com/story/news/local/sarasota/2018/03/12/bill-will-make-florida-most-transparent-state-in-nation/13004841007/>

<sup>6</sup> Laura Bischoff, *Ohio Court Data Project Promises to Bring Transparency, Improve Justice in Felony Cases*, THE COLUMBUS DISPATCH, OCT. 4, 2021, <https://www.dispatch.com/story/news/2021/10/04/ohio-sentencing-data-platform-tested-improve-court-transparency/5992267001/>

<sup>7</sup> Cook County State's Attorney, *Cook County State's Attorney Open Data Portal*, COOK COUNTY STATE'S ATTORNEY, 2022, <https://www.cookcountystatesattorney.org/about/open-data>

In this article, we look at one particularly rich and promising source of already legally public criminal data: criminal jury trial transcripts. We begin this article with a review of the legal limitations to public data access and with a discussion of what data is, practically speaking, public. We consider the current system of data access under a managerialized rights framework, where laypeople are required to take unjustifiably onerous steps to enact their rights. We then turn specifically to court transcripts as a rich data source demonstrating how courts generate and reify inequality, though these capacities of transcript data remain obscured by their inaccessibility. We then conduct an exploratory audit of court transcript data procurement policies across over 3,000 U.S. counties and draw thematic conclusions about de facto barriers to public access. We conclude with a discussion of how courts' approaches to public data need to change in the future, particularly in ways that lessen the burden on the public to access records they are legally entitled to.

## I. DATA RIGHTS VS. DATA ACCESS

### A. Legal Rights to Public Data

In the United States, the public has a right to access court proceedings and court records.<sup>8</sup> In *Nixon v. Warner Communications* the Supreme Court affirmed this right stating, “the courts of this country recognize a general right to inspect and copy public records and documents, including judicial records and documents (597).”<sup>9</sup> These rights have been reified consistently in courts across the country in the intervening decades. In the landmark case on the issue, *Richmond Newspapers, Inc. v. Virginia* (1980), the Supreme Court directly asserted that the

---

<sup>8</sup> See generally, Ronald D. May, *Public Access to Civil Court Records: A Common Law Approach*, 39 VANDERBILT LAW REVIEW 1465 (1986); James M. Chadwick, *Access to Electronic Court Records: An Outline of Issues and Legal Analysis*, U.S. DEPARTMENT OF JUSTICE, OFFICE OF JUSTICE PROGRAMS (for a detailed case history that enshrines these public rights and extends them to electronic versions of records).

<sup>9</sup> *Nixon v. Warner Communications*, 435 U.S. 589, 597 (1978).

First Amendment gives the public a constitutional right of access to criminal trials.<sup>10</sup> This was affirmed in *Globe Newspaper Company v. Superior Court* (1982), though both cases focused specifically on trial attendance rather than records.<sup>11</sup> Despite this focus, cases citing these decisions quickly expanded their focus to rights to access and examine court records, notably *Associated Press v. District Court* (1983) which held that the First Amendment required the court to provide pretrial records.<sup>12</sup> This is not to say that courts did not present any curtailing of these rights to access. Indeed, courts specified that lower courts had considerable discretion and advisory power over its own files and records.<sup>13</sup> This string of decisions, and debates surrounding the First Amendment, have continued until the present day, as legal questions surrounding access to electronic court records in particular have become more salient.<sup>14</sup>

#### B. Realities of Public Data Access

These legal protections for public access to criminal court data are staunchly juxtaposed with the realities of public data access. We must differentiate two types of public data to fully understand this dichotomy: *de jure* public data and *de facto* public data. Here we define *de jure* public data as data that is legally public. According to the law we cited above, vast swaths of criminal court data are *de jure* public data. The Administrative Office of the United States Courts clarify that there are some limitations to this overall legality: there are some situations in which data that would normally be public can legally be kept from the public. They describe the contours of these situations, saying,

---

<sup>10</sup> *Richmond Newspapers, Inc. v. Virginia*, 448 U.S. 555, 556-557 (1980).

<sup>11</sup> *Globe Newspaper Company v. Superior Court*, 457 U.S. 596, 601 (1982).

<sup>12</sup> *Associated Press v. District Court*, 705 F.2d 1143, 1148 (9th Cir. 1983)

<sup>13</sup> Ronald D. May, *Public Access to Civil Court Records: A Common Law Approach*, 39 VANDERBILT LAW REVIEW 1465, 1469 (1986).

<sup>14</sup> James M. Chadwick, *Access to Electronic Court Records: An Outline of Issues and Legal Analysis*. US Department of Justice, Office of Justice Programs 1, 1 (2001).

<https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/legal-issues.pdf>

“In certain circumstances, judges have the authority to seal additional documents or to close hearings that ordinarily would be public. Reasons can include protecting victims and cooperating informants and avoiding the release of information that might compromise an ongoing criminal investigation or a defendant’s due process rights.”<sup>15</sup>

This description further indicates that most criminal court data is not only occasionally public, rather it is ordinarily or presumptively public. Here we are not substantially concerned with the limited pool of data that is rarely defensibly and legally considered not public. This is a very small pool with exceptional constraints and to be overly focused on it at this analytic juncture would be to dismiss the larger systemic problem. Instead, we are precisely interested in the data that is ordinarily public. Even though this data is *de jure* public, and its public status is not technically fraught in any way, that does not mean it is actually practically possible for laypeople or researchers to obtain. It is worth noting that the case law begins to establish the contours of the right to access public records insofar as what documents and events are public and theoretically accessible but does not specify the mechanism through which that right can be enacted. Indeed, we argue that much of the data that is legally defined as public data is *de facto* non-public data because of substantial barriers to access. In this way we differentiate between public data and publicly accessible data by considering what these barriers to access are and how those barriers themselves contribute to enduring cycles of inequality in US court rooms.

We are not the first scholars to question the veracity and practicality of the term ‘public record.’ Particularly in the context of the tension between data privacy and data transparency, scholars have asked how public these public records really are for over two decades.<sup>16</sup> We extend

---

<sup>15</sup> United States Courts. *Accessing Court Documents – Journalists Guide*. UNITED STATES COURTS. Accessed February 15, 2022 at <https://www.uscourts.gov/statistics-reports/accessing-court-documents-journalists-guide>

<sup>16</sup> See Victoria Salzmann. *Are Public Records Really Public: The Collision Between the Right to Privacy and the Release of Public Court Records Over the Internet*, 52 BAYLOR L. REV. 355 for an overview of how FOIA, the right to privacy, and court decisions interact with electronic record rights and problems derivative of their disclosure.



this body of work by contemplating public data access using a framework of managerialized rights. Popular in studies of employment and more recently, Title IX policy, the managerialization of rights requires individuals to act on their own accord to secure protection of their rights from a larger institution.<sup>17</sup> This is because the rights exist but are not formally mechanized or handled by the grantor of the rights so must be managed by the grantee. In this way, protection or enactment of individual rights becomes a significant burden for the claimant rather than for the institution that is legally required to protect the rights of the claimant.

Scholars in this area clarify a number of stable features of managerialized rights processes, two of which are especially salient here. First, managerialized rights are often characterized by a group of people whose rights have been violated, but who are not vindicating their violated rights.<sup>18</sup> Second, these rights are not self-enforcing and are enforced differently depending on how the parties are related.<sup>19</sup>

---

<sup>17</sup> See generally work by Jeb Barnes & Thomas Burke, *HOW POLICY SHAPES POLITICS: RIGHTS, COURTS, LITIGATION, AND THE STRUGGLE OVER INJURY COMPENSATION* (2014); Ellen Berrey, *THE ENIGMA OF DIVERSITY: THE LANGUAGE OF RACE AND THE LIMITS OF RACIAL JUSTICE* (2015); Ellen Berrey, Robert L. Nelson & Laura Beth Nielsen, *RIGHTS ON TRIAL: HOW WORKPLACE DISCRIMINATION LAW PERPETUATES INEQUALITY* (2017); Thomas F. Burke, *LAWYERS, LAWSUITS, AND LEGAL RIGHTS* (2004); Frank Dobin and Alexandra Kalev, *Training Programs and Reporting Systems Won't End Sexual Harassment – Promoting More Women Will*, 70 *HARVARD BUSINESS REV.* 687 (2017); Frank Dobin and Alexandra Kalev, *Why Diversity Programs Fail*, 94 *HARVARD BUSINESS REV.* 14 (2016); LAUREN EDELMAN, *WORKING LAW: COURTS, CORPORATIONS, AND SYMBOLIC CIVIL RIGHTS* (2016); Lauren Edelman, *Legal Ambiguity and Symbolic Structures: Organizational Mediation of Civil Rights Law*, 97 *AMERICAN JOURNAL OF SOCIOLOGY* 1531 (1992).

<sup>18</sup> See generally work by Ellen Berrey, Robert L. Nelson & Laura Beth Nielsen, *RIGHTS ON TRIAL: HOW WORKPLACE DISCRIMINATION LAW PERPETUATES INEQUALITY* (2017); Kristin Bumiller, *THE CIVIL RIGHTS SOCIETY: THE SOCIAL CONSTRUCTION OF VICTIMS. REVISED EDITION* (1992); David Engel & Frank Munger, *RIGHTS OF INCLUSION* (2003). William L.F. Felstiner, Richard L. Abel & Austin Sarat, *The Emergence and Transformation of Disputes: Naming, Blaming, Claiming*, 15 *LAW & SOCIETY REVIEW* 631 (1980); Laura Beth Nielsen & Robert L. Nelson, *Rights Realized – An Empirical Analysis of Employment Discrimination Litigation as a Claiming System*, 2005 *WISCONSIN L. REV.* 663 (2005).

<sup>19</sup> See generally work by CATHERINE ALBISTON, *INSTITUTIONAL INEQUALITY AND THE MOBILIZATION OF THE FAMILY AND MEDICAL LEAVE ACT: RIGHTS ON LEAVE* (2010); David Engel, *The Oven Bird's Song: Insiders Outsiders, and Personal Injuries in an American Community*, 18 *LAW & SOCIETY REVIEW* 551 (1984).

Applied to the data access issues we discuss here; the public is legally entitled to public court records. However, the public is largely not vindicating their right to obtain these documents, which could be explained by several factors including a lack of knowledge of this right, a lack of understanding of opaque legal processes, a lack of detailed information necessary to initiate requests, or financial barriers to sustaining requests. This right to public records is also enforced differently across jurisdictions with different policies and is also differently accessible to different people. For instance, a law student might know more about contacting a county clerk than other people. Perhaps a private lawyer with a lucrative practice may be more able to financially afford records requests than other people, even more than other legal actors. In both hypotheticals, someone with less resources or knowledge who has an equal right to see data, data that may be the surest way to reveal patterns of harmful behavior, faces significant and unequal obstacles in vindicating their rights. Managerialized rights are therefore an access to justice problem because they shift the burden for protection to the exploited party, who is often powerless in the face of the larger institution.

This conceptualization is a useful one for considering the relationship between the public and supposedly public data because it helps interpret barriers to public access. Despite court records being *de jure* public, the current structure places the onus almost entirely on the requestor. The process by which court records are requested requires that individuals who seek court records have specialized knowledge of where to obtain the records, have specialized knowledge about the individual cases that they seek records of, that they have sufficient financial resources to procure the records, and that they can follow whatever local procedure the court deems necessary to obtain those records. Individuals have theoretical legal access to these records, but substantial practical barriers prevent actual access.

Take for example, the “Public Access to Court Electronic Records” or PACER system maintained by the federal judiciary. This system touts itself as a public records service that provides the public with “instantaneous access to more than 1 billion documents filed at all federal courts.”<sup>20</sup> However, this access comes at a price. There are two major roadblocks in this system: first, the logistical and procedural difficulty of PACER and, second, the cost of accessing records. First, an interested user has to be aware of PACER and must know if the types of documents they seek are available through PACER at all. They must also request an account on the PACER website and wait several business days for their log-in information to be mailed to a physical address. Furthermore, an interested user has to have enough information about a given case to motivate a search through the PACER portal which may include information outside of the experience of the lay user, such as docket numbers.

Second, and relatedly, the interested user has to be financially able to see their request through. PACER charges a minimum \$0.10 per page, including calculating bill-able pages out of HTML-formatted information. This may not sound like a lot, but PACER builds in charges that functionally penalize users who are less knowledgeable. If, for example, a user was to enter a party name that yielded multiple matches, they would be charged for every resultant match, not just for the data that they actually wanted.<sup>21</sup> In this way, the court has continued to endorse a fee structure that makes public court records practically inaccessible to most of the public under the guise of increased accessibility. Notably, this does not serve to protect individuals’ personal information necessarily, as personal information has been monetized by companies who are

---

<sup>20</sup> PACER, The Public Access to Court Electronic records FAQ, UNITED STATES COURTS.  
<https://pacer.uscourts.gov/>

<sup>21</sup> PACER, The Public Access to Court Electronic records FAQ, UNITED STATES COURTS.  
<https://pacer.uscourts.gov/>

willing to share data the court obscures – for a fee.<sup>22</sup> As we will discuss, this problem extends beyond PACER and permeates the US criminal legal system more broadly.

The courts seem to provide additional access options to these data for free, perhaps suggesting a commitment to public access. The website for the Administrative Office of the US Courts says, “Electronic records can be viewed in the clerk of court’s office for free, as can any paper records that have not been destroyed or transferred to the National Archives. But per-page fees are charged for printing or copying court documents in the clerk’s office”.<sup>23</sup> In this version of public access, a member of the public must secure both time and transport to the clerk of court’s office and they must hope that the documents they want have not been destroyed. If those documents do exist, they must still pay per page for copies of them, which does not function to alleviate the burdens to public that already exist through the online system.

Therefore, much like other examples of managerialized rights, a member of the general public who wishes to exercise their rights to see court records is responsible for taking on the onus of the request, for procuring specialized knowledge about the request, and financially sponsoring their request even in error. Notably, while these per-page fees might seem small if executed correctly, court records are notoriously lengthy. This fee structure then makes it nearly impossible to collect court records at scale necessary to audit various elements of the justice system even if the expert knowledge to complete a mass-scale request is known. The Systematic Content Analysis of Litigation EventS Open Knowledge Network (SCALES OKN), a group who aims to procure and disseminate public record records, estimate that procuring the documents for

---

<sup>22</sup> Arminda Bradford Bepko, *Public Availability or Practical Obscurity: The Debate Over Public Access to Court Records on the Internet*, 49 *NYL SCH. L. REV.* 967, 982 (2004).

<sup>23</sup> United States Courts. *Accessing Court Documents – Journalists Guide*. United States Courts. Accessed February 15, 2022 at <https://www.uscourts.gov/statistics-reports/accessing-court-documents-journalists-guide>

a single year of partial data cost them over 100,000 dollars.<sup>24</sup> This further demonstrates the reality that *de jure* public records are actually *de facto* non-public records.

## II. COURT TRANSCRIPTS AS PUBLIC DATA

We turn now to an analysis of one particular type of public data, criminal jury trial transcripts, that is particularly disadvantaged by the current system of managerialized record retrieval. Court transcripts are unlike other types of court data in that they capture rich rhetorical data about court proceedings and contain record of various process actions that are not recorded in other types of court documents.<sup>25</sup> They also contain and can be mined for basic procedural and substantive data that is not necessarily otherwise collected or distributed. The outcome of this data processing is plausibly datasets of information, some of it very basic, that is otherwise not aggregated. They also vary substantially from other types of court record by being indeterminately lengthy and virtually excluded from even the paid public records services currently offered by the many courts. We are not the first to encounter this problem. Notably, other scholars study the universe of criminal record and criminal court data specifically point to transcripts as some of the most logistically and financially difficult court records to obtain.<sup>26</sup>

### A. Case Study: Cook County, Illinois

Explaining the trouble of procuring criminal jury transcripts is most easily done via example, so here we use the example of Cook County, Illinois to exemplify the range of access

---

<sup>24</sup> Pah, Adam R., David L. Schwartz, Sarath Sanga, Zachary D. Clopton, Peter DiCola, Rachel Davis Mersey, Charlotte S. Alexander, Kristian J. Hammond, and Luis A. Nunes Amaral. 369 *How to Build a More Open Justice System*, SCIENCE 136 (2020)

<sup>25</sup> See David DeMatteo, John F. Edens, Meghann Galloway, Jennifer Cox, Shannon Toney Smith, Julie Present Koller, & Benjamin Bersoff, *Investigating the Role of the Psychopathy Checklist-Revised in United States Case Law*. 20 PSYCHOLOGY, PUBLIC POLICY, AND LAW 96, 105 (2014) (who argue that trial transcripts would allow researchers access to study cases that do not result in written opinions).

<sup>26</sup> David Bright, Russell Brewer, and Carlo Morselli, *Using Social Network Analysis to Study Crime: Navigating the Challenges of Criminal Justice Records*, 66 SOCIAL NETWORKS, 50, 54 (2021).

to justice problems with court transcripts. As we move through this example, it should be noted that Cook County varies from many of its contemporaries by actually having a policy by which one can clearly request court transcripts. The difficulty in obtaining transcripts starts with the reality that even basic facts about the workings of the criminal legal system are opaque. This problem is so egregious, that even the number of criminal trials in Cook County is not readily communicated to the public. The Administrative Office of Illinois Courts Annual Report states that of the small percentage of cases that went to trial, only 10% of those went to a jury trial while the other 90% went to a bench trial.<sup>27</sup> It is much more difficult to find out how what the raw number of such cases generally is, much less which cases they are specifically. This becomes crucial because of the burden placed on the public when requesting the court transcripts that we argue are uniquely necessary to reveal the frequency of legal events like racially motivated peremptory strikes.

Taking on the role of a member of the public who wishes to procure a court transcript, you might begin by trying to Google something like ‘cook county court transcript.’ The first result will conveniently take you to the Cook County Clerk’s page where you will be greeted by an empty webpage containing two clickable buttons, captioned “Contact Information” and “Transcript Orders & Rates,” both of which are revealed to do absolutely nothing (see Figure 1).<sup>28</sup>

---

<sup>27</sup> Circuit Court of Cook County, *Trials and Other Criminal Proceedings in Cook County have Continued Through Pandemic*, STATE OF ILLINOIS, Ap. 20, 2021 <https://www.cookcountycourt.org/MEDIA/View-Press-Release/ArticleId/2836/Trials-and-other-criminal-proceedings-in-Cook-County-have-continued-through-pandemic> Here we cite a 2019 statistic, before COVID-19 related functioning disruptions given by the Circuit Court of Cook County. In the same release, they state that nationally 92% of felony cases are resolved via dismissal or guilty pleas and imply Cook County is similar, but do not state so explicitly.

<sup>28</sup> As of February 16, 2022. See <https://www.cookcountycourt.org/ABOUT-THE-COURT/Office-of-the-Chief-Judge/Court-Related-Services/Official-Court-Reporters/Contact-Information>.

Figure 1: Cook County Clerk Transcript Orders and Rates



If this result does not discourage you and you go back to Google, adding terms and clicking on seemingly less-related webpages you may click on a link marked ‘Court Reporters’ that is revealed to have transcript information at the bottom.<sup>29</sup> These tabs will reveal what types of information you must have to successfully request a single trial transcript. You may be asked for the date of the hearing, the name of the case, the case number, the branch or courtroom number where the case was heard, and the name of the judge.

Assuming you have all of this information, the financial considerations begin. If you opt for copies of the transcript rather than originals, you will be required to pay 2.00 per page, for transcripts that can easily exceed 100 pages. There is also a tab that touts itself as a guide to instruct you where to go to order the transcript. This guide requires you to know where each individual case was heard, understand the legal terminology, and scroll to your destination. The sum of this exercise, then, is to make it clear that even requesting one single transcript is

---

<sup>29</sup> Cook County Courts, *Court Reporters*, Accessed February 16, 2022 at <https://www.cookcountycourt.org/ABOUT-THE-COURT/Office-of-the-Chief-Judge/Court-Related-Services/Official-Court-Reporters>

confusing, is expensive, and is obfuscated by access issues on the part of the courts. This labyrinthian process for obtaining a single transcript is not scalable. It becomes even more implausible to consider that a sufficient number of transcripts for computational analysis could be obtained through this process with any sort of efficiency. This means that a relatively simple computational project, perhaps requiring only basic natural language processing, is stymied by court processes that turn accessing legally public records into an insurmountable barrier.

Perhaps the most sobering part of this reality, is that Cook County stands above many of its peers by having a specific and public-facing transcript policy at all. Though unacceptably labyrinthian, Cook County's process is uniquely transparent and easy to navigate. As we will discuss, the difficulty is substantially higher in other counties across the US. Much like the managerialized rights framework with which we considered public data access above, the burden on a member of the general public to enact their right to inspect public criminal trial transcripts is high, even in a location like Cook County which offers some form of process and some public information about that process. As we reveal in this Article, public information about transcript retrieval processes is rare, further compounding the access problems to the records and our ability to study justice problems uniquely discernable from criminal jury transcripts.

Consequently, there is limited research that has been undertaken using court transcripts and what research has been done is often limited to particular cases or small sample sizes. Among the larger US transcript sample sizes found by these authors was work by Boothroyd et al. (2003) who analyzed 104 transcripts, Belli et al. (2015) who analyzed 15 transcripts, and Hoppe (2014) who analyzed 43 transcripts.<sup>30</sup> While these studies make good use of the transcript data, it is

---

<sup>30</sup> Boothroyd, Roger A., Norman G. Poythress, Annette McGaha, and John Petrila. 2003. "The Broward Mental Health Court: Process, Outcomes, and Service Utilization." *International Journal of Law and Psychiatry* 26(1):55, 57; Belli, Roberta, Joshua D. Freilich, Steven M. Chermak, and Katharine A. Boyd. 2015. "Exploring the Crime-Terror Nexus in the United States: A Social Network Analysis of a Hezbollah Network Involved in Trade



clear that the stopgap preventing further analysis of court transcripts is not manual efficiency, much less computational efficiency, but is rather access to supposedly public documents. With increased ability to request and obtain court transcripts, we can exponentially increase the universe of potential analysis and make use of new computational legal techniques to study the contents of those transcripts, analyzing thousands or tens of thousands, rather than dozens, at a time. This would allow researchers to analyze the entire data universe, rather than sectioning out small portions of it.

#### B. Current Limitations to Knowledge Due to Transcript Inaccessibility

Access to criminal court transcripts would substantially alter the terrain of what is known about criminal courts, both for interested researchers and the general public who seek to understand the functioning and context of courts that enact law upon them. Specifically, access to transcripts for criminal jury trials would grant access to two different types of information: information on the basic functioning of the court and information on access to justice and equity issues within the court. Although this paper is keenly concerned about the latter, with particular attention to issues of equity with respect to jury selection, it is worth discussing the potential insight into the basic functioning of the court and how that, in itself, can be a question of access and equity. We spend considerable time here considering how transcripts could be particularly advantageous not to privilege transcripts above other types of currently inaccessible records, but rather to demonstrate how these types of inaccessible records becoming available might alter the

---

Diversion." *Dynamics of Asymmetric Conflict* 8(3): 263, 266; Hoppe, Trevor. 2014. "From Sickness to Badness: The Criminalization of HIV in Michigan." *Social Science & Medicine* 101:139, 143.

terrain of justice. This is a non-exhaustive account of some of the ways in which access to court transcripts could improve knowledge about the courts.

### *1. Court Functioning*

First, we look into insights that transcripts can provide on how the court functions. As we discuss in this section, broad insight into patterns and practices of the courts as a system is difficult to obtain. Researchers and policymakers are often limited by time consuming methodology that is prohibitive to scale for even single major metropolitan areas: practitioner interviews and court watching make up most of our current knowledge landscape about the day-to-day functioning of the courts. We offer several ways in which access to criminal court transcripts could enhance understanding.

The first area in which transcripts would provide insight into court functioning is in the use and presentation of evidence. The presentation of evidence obviously has substantial relevance for appeal. The improper presentation of evidence in a criminal case can itself be grounds for appeal.<sup>31</sup> The use of evidence, in other words, is integral to understanding the routine function of the court and the potential stability of convictions and acquittals.

Transcripts would offer full insight into when and how evidence is presented as well as how objections are sustained or overruled. In the jury context, this information would give particular insight into the use of the rhetorical presentation of technically inadmissible evidence: on the frequency with which juries are asked to dismiss the presentation of evidence following a sustained objection. This has substantial implications, then, for understanding courtroom norms on a broad level.

---

<sup>31</sup> This can involve counsel's failure to present evidence, the prosecution's failure to share potentially exculpatory evidence (under *Brady v. Maryland*, 373 U.S. 83 (1963)), or the improper consideration of inadmissible evidence.

Methodologically, we are currently able to gain some understanding of the use and presentation of evidence through practitioner interviews and court watching but transcripts would offer a more robust account. Interviews allow for the collection of practitioner perspectives about evidence and court watching provides a snapshot of particular individual cases but neither offers a robust systemic analysis. Furthermore, a great deal of available contemporary scholarship on the presentation of evidence before juries center around a relatively narrow area of evidence: forensic evidence.<sup>32</sup>

Transcripts would also aid in understanding the general rhetorical presentation of the case to a jury. Currently, we lack a complete account of, for example, the modes of argumentation being made to jurors in opening or closing arguments. The totality of our understanding of this genre is more fully fleshed out from courtroom dramas than from primary source material from the courts. Again, insight into rhetorical techniques can currently be gleaned in part through practitioner interviews and court observations. However, practitioner interviews can truly only give insight into how practitioners conceptualize their own rhetorical techniques, not how they're presented to an audience or interpretable in context. Court observations are limited in their overall scope: we're unable to truly see the court as a system through this limited mode of data

---

<sup>32</sup> Here we refer in particular to the well-established literature about the CSI Effect, or the highly debated theory that jurors are unduly influenced by fictional crime television media such that it has meaningfully changed their expectations of forensic evidence *see e.g.* Steven M. Smith, Veronica Stinson & Marc W. Patry, *Fact of Fiction: The Myth and Reality of the CSI Effect*, 4 CT. REV. 1 (2011); Kimberlianne Podlas, *The CSI Effect and Other Forensic Fictions*, 27 LOY. LA. ENT. L. REV. 87 (2006); Simon A. Cole & Rachel Dioso-Villa, *CSI and Its Effects: Media, Juries, and the Burden of Proof*, 41 NEW ENG. L. REV. 435 (2007); Simon A. Cole, *A Surfeit of Science: The "CSI Effect" and the Media Appropriation of the Public Understanding of Science*, 24 PUB. UNDERSTANDING SCIENCE 130 (2013); Andrew P. Thomas, *The CSI Effect: Fact or Fiction*, 115 YALE L. J. POCKET PART 70 (2006), <https://www.yalelawjournal.org/forum/the-csi-effect-fact-or-fiction>; Donald E. Shelton, *The 'CSI Effect': Does It Really Exist?*, 259 NAT'L INST. JUST. J. 1 (2008); Young S. Kim, Gregg Barak, Donald E. Shelton, *Examining the "CSI-effect" in the Cases of Circumstantial Evidence and Eyewitness Testimony: Multivariate and Path Analyses*, 37 J. CRIM. JUST. 452 (2009); *State v. Cooke*, 914 A.2d 1078 (Del. 2007); Nicholas J. Schweitzer & Michael J. Saks, *The CSI Effect: Popular Fiction About Forensic Science Affects the Public's Expectations About Real Forensic Science*, 47 JURIMETRICS 357 (2007) and others.

collection. Further, researchers should not need to expend additional resources when this information is theoretically available via public documents.

A rhetorical analysis of opening and closing arguments would allow for an understanding of what types of appeals practitioners present as effective for juries or what the universe of arguments they present as available for presenting to juries. This tells us not what juries do or do not respond to but what practicing lawyers actually use in practice, presumably under the assumption that it would be persuasive.

This would also allow for an understanding of how practitioners present technical or specific legal jargon. How do they unpack what has happened, or what is about to happen, in layman's terms for a jury that is unlikely to be made up of fellow practitioners? What types of work is being done to allow these particular factfinders to make accurate legal decisions – and how might that interplay with certain strategic considerations? The answers to both of these questions involve the rhetorical presentation to the jury that have substantial implications for justice. The presentation of technical information speaks to the jury's fundamental understanding of their jobs and whether or not they are able to do it effectively.

Finally, a complete set of transcripts would allow for a fulsome understanding of delays in specifically analyzing when and how continuances are issued. Currently, we're able to understand some of this through practitioner and stakeholder interviews as well as through court observations.<sup>33</sup> Again, what we lack is a systemic analysis of when and where continuances occur. Being able to analyze a complete universe of data to point to patterns in moments – if

---

<sup>33</sup> Maria Hawilo, Kat Albrecht, Meredith Martin Rountree, and Thomas Geraghty, *How Culture Impacts Courtrooms*, 112 THE JOURNAL OF CRIMINAL LAW AND CRIMINOLOGY 171, 188-189; Kat Albrecht, Maria Hawilo, Thomas Geraghty, and Meredith Martin Rountree, *Justice Delayed: The Complex System of Delays in Criminal Court*, LOYOLA UNIVERSITY OF CHICAGO LAW SCHOOL [currently in production, will have page and volume numbers shortly]

they always happen because one party is not present or because paperwork has not been properly transferred between divisions or departments – allows for a better understanding of systemic failures.

Although the use of continuances is a regular court functioning, the issue of delays in the court is a matter of equity and access to courts.<sup>34</sup> Continuances or delays mean more court dates which has the snowballing effect of increased logistical issues for litigants, witnesses, or defendants: taking (more) time off from work, finding (more) childcare, and obtaining (more) transportation to and from the courthouse. Furthermore, delays can extend time to trial for criminal defendants, having the impact of keeping people incarcerated for longer while awaiting trial.

## 2. *Access and Equity Issues*

Additionally, a lack of transparent court data has direct implications for access and equity issues by substantially limiting the types of work that can be done in several key areas.

Although, as we mentioned, some jurisdictions are becoming more public with some of their data, a good deal of knowledge here is currently lacking even in jurisdictions with more robust public data. Key examples of this are in the difficulty of studying judicial temperament and demeanor, the difficulty in robustly studying jury instructions (and whether they are understood by the jury), and, in the key case study in this paper, in understanding racial discrimination in jury selection.

First, information on judicial temperament and demeanor is difficult to ascertain without complete access to transcripts. Currently, our best understanding of how courtrooms are

---

<sup>34</sup> Kat Albrecht, Maria Hawilo, Thomas Geraghty, and Meredith Martin Rountree, *Justice Delayed: The Complex System of Delays in Criminal Court*, LOYOLA UNIVERSITY OF CHICAGO LAW SCHOOL [currently in production, will have page and volume numbers shortly]

managed is assessed through court observations and interviews with practitioners. Although scholarship is not in the business of evaluating judges for competency, understanding how judges manage their courtrooms speaks generally to how litigants, defendants, victims, and witnesses are treated and can give substantial insight into where and why systemic issues arise. In other words, a robust evaluation of the systemic failures of the courts must necessarily include an evaluation of how judges manage those spaces and the treatment of the people before them.

Second, transcripts could plausibly provide a universe of data on jury instructions and their comprehensibility for the juries. Although questions from jurors pertaining to clarifying their instructions or role do not necessarily go back on the record, this is an area that could conceivably show up on transcripts. Without, first, a full collection of public transcripts, it is difficult to even ascertain the rate at which such questions go on record. The questions themselves are unlikely to be fully answered, given the overall judicial reticence to do so, having potential access to the universe of questions that are asked repetitively can provide insight into areas where those particular instructions fail to actually *instruct* juries.

Finally, having a comprehensive set of data on the *voir dire* portion of criminal jury trials would allow for a comprehensive understanding of jury selection – specifically the use of racially motivated peremptory strikes – post-*Batson*. We discuss this example in some depth because it is the only such example whereby the public – through journalists – had unique access to a large volume of criminal trial transcripts. In 1986, the United States Supreme Court held in *Batson v. Kentucky* that the prosecutor could not strike Black jurors on the basis of race, without providing a neutral reason for the strike.<sup>35</sup> They held that to do so would undermine the public perception of fairness in the criminal legal system and would deprive defendants of basic

---

<sup>35</sup> *Batson v. Kentucky*, 476 U.S. 79, 97 (1986).

rights.<sup>36</sup> Since that decision, the Supreme Court has provided some clarification on how to determine if the prosecution has given an adequately and honestly neutral alternative explanation (“legitimate nondiscriminatory reason”) for their striking of Black jurors. In other words, the Supreme Court has since begun to weigh in on how a trial judge might discern between a legitimate nondiscriminatory reason for striking a juror and a reason given as a proxy for discrimination against non-white jurors – and what a judge could legally consider in evaluating that distinction.

Prior to 2019, the ability to argue that behavior was, in fact, a proxy for discrimination was exceptionally thin. The Supreme Court ruled in 2005 in *Miller-El v. Dretke* that a defendant could rely on “all relevant circumstances” in considering purposeful discrimination.<sup>37</sup> Here, the relevant circumstances were a compilation of jury shuffling, disparate patterns of jury questioning by race, prosecutor notes, and the statistical fact that the state issued a pre-emptory challenge to only 12% of nonblack potential juries, while eliminating 91% of potential black jurors.<sup>38</sup> In 2008, the Supreme Court extended this in *Snyder v. Louisiana* by ruling that it is... “[E]nough to recognize that a peremptory strike shown to have been motivated in substantial part by discriminatory intent could not be sustained based on any lesser showing by the prosecution.”<sup>39</sup> Similarly, the Supreme Court held in 2015 in *Foster v. Chatman* that the prosecutor’s notes, disqualifying Black jurors prior to questioning, was sufficient to establish the intentionality.<sup>40</sup> Further, the Court then pointed to the presence of similarly situated white jurors who were not struck.<sup>41</sup>

---

<sup>36</sup> *Batson v. Kentucky*, 476 U.S. 79, 87 (1986).

<sup>37</sup> *Miller-El v. Dretke*, 545 U.S. 231, 240 (2005).

<sup>38</sup> *Miller-El v. Dretke*, 545 U.S. 231, 265-266 (2005).

<sup>39</sup> *Snyder v. Louisiana*, 552 U.S. 472, 485 (2008).

<sup>40</sup> *Foster v. Chatman*, 578 U.S. \_\_\_, 501 (2016).

<sup>41</sup> *Foster v. Chatman*, 578 U.S. \_\_\_, 511 (2016).

In 2019, in *Flowers v. Mississippi*, the Supreme Court consolidated and affirmed these markers of proxy and extended them. The Court held that the State’s apparent relentlessness in pursuing an all-white jury, the State’s additional time spent questioning Black perspective jurors (compared to accepted white jurors), and the difference between the Black struck jurors and white jurors established that the State’s given reasons for striking Black jurors were proxy for racial discrimination.<sup>42</sup> Finally, and most importantly, the Court considered that the State’s use of peremptory strikes in the trial in question (Flowers’s sixth trial) followed the same pattern as in the first four trials (two of which had resulted in a conviction being overturned for prosecutorial misconduct in wrongfully striking Black jurors). In other words, the Supreme Court has plausibly opened the door for the consideration of patterns of racial discrimination in the use of peremptory strikes, as long as that can be established through data.

*Flowers v. Mississippi* was an incredibly unique case in the *Batson* lineage for two major reasons: first, the facts of *Flowers* were extraordinary and, second, the court had unique, unprecedented access to the prosecutor’s entire history of peremptory strikes due to independent data collection and analysis done by reporters for American Public Media.<sup>43</sup> *Flowers* affirmed that the trial judge can consider statistical evidence of the State’s use of peremptory strikes against Black jurors versus white jurors in a case as well as considering that prosecutor’s relevant history in past cases.<sup>44</sup> To this point, the Supreme Court has also affirmed that this evidence is not necessary – that the defense need not establish a pattern of discriminatory

---

<sup>42</sup> *Flowers v. Mississippi*, 139 S. Ct. 2228, 2248 (2019).

<sup>43</sup> APM Reports produces a podcast entitled *In the Dark*. The second season covered the Curtis Flowers case and, in doing so, the reporters physically combed through the available transcripts of all of prosecutor Doug Evans’s criminal jury trials in his career. This is currently the most robust investigation into criminal legal transcripts and it was acquired via trips to the courthouse and analyzed manually. Their transcript collection is here: <https://features.apmreports.org/in-the-dark/season-two/source-notes.html>

<sup>44</sup> *Flowers v. Mississippi*, 139 S. Ct. 2228, 2243 (2019).



behavior – however, in this case, that pattern of behavior in the history of the Flowers trial was available.<sup>45</sup>

As mentioned, the *Flowers* case was unique in that Curtis Flowers was tried six times for the same murders.<sup>46</sup> Although the Supreme Court limits their analysis of a historical pattern to Evans’s behavior only in prior Flowers cases, they decidedly hold that the trial judge can consider context outside of the unique trial at hand. Conceivably the Court could limit court pattern analysis to cases with multiple trials against the same criminal defendant on that same issue. However, the Court does not make that limitation explicit within the opinion. Regardless, it is worth noting that such information is extraordinarily difficult to come by and the door opened by the Court’s allowance of the consideration of context is potentially impossible to get through for reasons explained elsewhere in this paper. Therefore, the *Flowers* case serves as an exemplary case study of how systemic discrimination within the legal system can be better understood – and better argued before the courts – due to increased access to transcripts.

### III. AUDIT OF U.S. TRANSCRIPT AVAILABILITY

Our goal in this article is not just to indict the lack of available data and call for more, but we also aim to analyze how the data that is available continues to shift the onus for procuring theoretically public documents to a public that cannot meaningfully access them. Therefore, we supplement our analysis of this data universe with an exploratory audit of transcript availability and record procurement policies across the United States. Broadly speaking, we seek to analyze

---

<sup>45</sup> *Swain v. Alabama*, 380 U.S. 202 (1965) establishes that requiring a pattern is too strict, but *Batson* allows for the consideration of a pattern.

<sup>46</sup> *Flowers v. Mississippi*, 139 S. Ct. 2228, 2234 (2019).

exactly how inaccessible court transcripts are by emulating the process that laypersons are expected to follow in order to obtain transcripts from their local courthouses.

Notably, we follow this process in the way that laypersons would rather than researchers, who may have specific connections to courts or financial resources unobtainable to the general public. We do so intentionally because using well-resourced academics as a benchmark for public access violates the spirit of the law surrounding public access. That is, academic researchers are not sole representatives of the ‘public’ that is legally entitled to access public records. That said, the research team that coded these policies are still likely more knowledgeable about legal systems and information location than the average layperson. We address this reality in the discussion of our results at length but emphasize this point here to contextualize the methodological processes followed our audit of transcript availability.

Our goal in conducting this audit is to understand inaccessibility as a multi-dimensional concept. Something can be inaccessible for a number of reasons. For example, it may be inaccessible by simply not existing, a relatively straightforward definition of inaccessibility. However, something might also be inaccessible because barriers to access may be practically insurmountable for the target audience. Here we consider not just the existence of information on procuring criminal court transcripts, but also what type of information about a legal case you need to have to request the transcript, how much the transcripts will actually cost, and the actual process for literally obtaining the transcript after making your request.

#### A. Methodology

We conducted the transcript audit at the county level, using Census Areas, districts, and boroughs where appropriate.<sup>47</sup> We opted to do so for several methodological reasons: 1) laypersons generally encounter legal processes on a closer-to-local level rather than at the state level, 2) if there is consistency in process at a higher unit of analysis, that will become apparent at a finer grained unit of analysis rather than artificially reducing variation out of the sample, 3) to capture a full range of variation within legally cognizable jurisdictions. Our developed list of areas that meet our inclusion criteria included 3,244 valid areas, a vast majority being counties in the contiguous 48 states of the United States.

Each county was coded manually, by the authors of this manuscript. Because of their positionality as legal academics and practitioners, the coding team was aware that court record policies (that would include transcripts) are generally under the purview of the county clerk. However, in order to replicate the process that would be followed by a layperson, the coders instead initiated standard searches in a search engine to find the information using terms like ‘[county name] transcript court records’ and ‘[county name] records request.’

Following our theory of managerialized rights and the specific burdens placed on the general public to obtain transcripts, we code for a variety of *de facto* barriers to public access. First, we include a blanket code for whether or not information on transcripts was available at all. Second, we record information about the price of transcripts, if it is given. Third, we record any information about what specialized information a requestor needs to request a transcript. Fourth, we record practical logistical details about an individual would actually obtain the transcript

---

<sup>47</sup> We include boroughs, Census Areas, and districts in the 50 United States, unincorporated territories of the United States, and territories of the United States because the US exercises sovereign powers over all such areas.

(both in process and in mode of request/delivery). We do this for each of the identified 3,244 valid areas.

## B. Results

The results of this audit reveal numerous barriers to public access for legally declared public data. We find these access barriers at virtually every coded nexus of possible variation. We discuss in this article several salient themes that emerged across the data distribution. Here we focus not on the fine-grained analytics of the audit sample, instead opting to consider how these thematic categories relate to the legal issues dealt with in this Article.

### *1. Missingness*

The first and most pervasive theme of this analysis was missingness. It was common throughout the data for counties to simply have no locatable public-facing policy on how to obtain court records at all. In these cases, the researchers searched exhaustively for record request policies of any kind but were often unable to locate any at all. What we encountered instead were a number of websites that claimed to have county level records, despite being privately held.<sup>48</sup> This poses a problem to the general public who may be relying on these sights when official sources are not present. Even in cases where these websites are well-intentioned, they do not constitute a sufficient solution to protecting public data access, when that burden is rightly ascribed to the courts.

It was also common that counties would have pages instructing how to obtain court records but would specify and limit those types of records in ways that did not seem to include

---

<sup>48</sup> A typical example of this is <https://countyclerkrecords.com/>, who note in their disclaimer at the bottom of the page that “CountyClerkRecords.com is a privately owned informational website that is not owned or operated by any state government agency.” These sites generally are repositories of information relating to numerous geographies and appear in search readily when official websites do not exist.

transcripts. In fact, court transcripts were very rarely mentioned at all, despite having substantially different logistical requirements than other types of public records. Qualitatively, this was especially common in smaller counties, whose data policies almost exclusively referred to property data. Larger and more traditionally urban counties were more reliable in including reference to transcripts specifically in their data policies, though explicit reference to transcripts remained abnormal across the data.

The simple presence of information was not the only type of missingness. Also frequent in the data were obfuscations about general process or a lack of detail that functionally equivocated to missingness. For example, in Benton County Arkansas, there was some information about court records, but the website instructed that you must call a phone number for more information.<sup>49</sup> We conceptualize this as a barrier to access because it requires additional agency to be expended by the requestor rather than communicating that information as transparently as possible to a potential requestor.

## *2. Procedural Variation in a Legally Identical Process*

The second theme that we noticed throughout the data was highly varied contact procedure for what should be a legally identical process. That is, it is the job of county clerks to act as stewards and disseminators of public records. However, this is not consistently the process by which counties are instructing individuals to obtain records. Instead, processes varied substantially in how likely it would be for a layperson to identify the proper recipient.

In some cases, jurisdictions had portal-like systems allowing a requestor to fill in a contact sheet to make a data request. One example of this was Barrow County, Georgia who pointed

---

<sup>49</sup> Benton County, Arkansas Policy - <https://bentoncountyar.gov/circuit-clerk/public-records/>

requestors to such a portal, but again, did not specifically describe transcript procurement.<sup>50</sup>

Other jurisdictions required significantly more specificity in identifying the proper contact person to obtain a transcript. For example, in Arenac County, Michigan requestors must contact the specific court reporter who was assigned to the relevant judge who heard the case.<sup>51</sup> This is another example of increasing burdens to public access. In general, we note three ways of dealing with court record requests: 1) the County Clerk handles the request, 2) the requestor must contact a court reporter directly, 3) a third party entirely handles the request.<sup>52</sup>

We found this variation significant not just in its range of substantive difficulty, but in how it speaks to system-wide dysfunction. Public right to public data is not a local-level protection, that some counties can opt into and other can opt out of at will. Rather what we identified is procedural variation in what should be a legally identical process across jurisdictions. Having such widely varying processes makes accessing records in certain jurisdictions virtually impossible, when the same record can be much more easily obtained in a different location.

### *3. Financial Burdens and Outdated Processes*

The third salient theme of this analysis concerns financial burdens and outdated processes that prevent public access to public data. Information about prices specific to transcript requests were rare. Across counties and states per page pricing of transcripts ranged from 0.25 per page to 2 dollars per page, when information was given. Much more common was a lack of

---

<sup>50</sup> <https://www.barrowga.org/departments/firm-record-request.aspx>

<sup>51</sup> <https://www.arenaccountymi.gov/Courts-Law/23rd-Circuit-Court/>

<sup>52</sup> An example of latter case, where a third party entirely handles transcript requests is Jackson County, Michigan. Jackson County refers transcript requests to ‘Theresa’s Transcript Service,’ again emphasizing that despite consistency in legal entitlements to public data, the process of getting it is highly variable and often under the purview of unexpected actors. (<https://www.co.jackson.mi.us/443/Transcripts>)

information about what to expect to pay for a criminal jury trial transcript, since such documents were not mentioned at all.

However, there were also other additional fees attached to record retrieval including research fees, shipping and handling fees, and service fees that were more commonly quoted in county policies. Most commonly assessed were research fees, which ranged from 20-30 dollars. In some counties, like Pima County, Arizona, a requestor would be assessed a 30-dollar fee for every name that had to be searched by the Clerk, meaning that research fees could become quite substantial for one request<sup>53</sup>. The policies also stated that requestors should expect to often pay shipping and handling fees for physical copies of documents to be sent to their location. The price of these fees, often around 7 dollars, clearly indicate that Clerks are not anticipating sending bulky criminal jury trial transcripts. Some counties, like La Paz County, Arizona, require requestors to send them a self-addressed and stamped envelope to receive their documents.<sup>54</sup> Not only is this a slow method of delivery that requires additional effort on the part of the requestor for no discernable reason, it also delineates the expectations of the courts that record requests should be able to fit in an envelope. Easily exceeding hundreds of pages in length, there is no plausible envelope that could hold a single criminal jury trial transcript. Some counties did have online options for document procurement and request, while others still principally recommended mail, fax, or even in-person pick-up.

#### *4. Expert Knowledge Requirements*

The fourth and final theme we explore here concerns the amount of information required to motivate a search for records. Counties generally had high standards for the amount of

---

<sup>53</sup> <https://www.jp.pima.gov/Forms/JP90C%20-%20Records%20Fee%20Schedule%20%208-9-17.pdf>

<sup>54</sup> <http://www.co.la-paz.az.us/DocumentCenter/View/4200/Fee-Schedule-PDF>

information required to motivate a search, should you not want to incur additional research fees payable to the Clerk. At minimum these generally included case number and party name, but requests for information also might include the specific date, the name of the judge, and the number of pages in the document. Requiring this level of information about each request 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. Generally, these criteria were not consistent across states or counties, which would further amplify the difficulty in making either individual or mass requests.

Some counties required substantially less information to process a search, indicating that searches are possible to enact with much more limited information, at least in some cases. This is an encouraging sign that some of these expert-level knowledge requirements could be changed, even under the existing framework. Cochise County, Arizona requires at minimum only the defendant's name and the approximate date, while Autauga County, Alabama asks only for a 'reference' or 'description' of the case.<sup>55</sup>

### C. Discussion

The exploratory results of our transcript process audit are consistent with a systemic process of distancing governmental responsibility and managerializing public access to public records. Across all four themes we describe here: missing information, varied contact procedure, financial burdens and process inefficiency, and expert knowledge requirements on the part of

---

<sup>55</sup> <https://www.cochise.az.gov/191/Requesting-Copies-of-Court-Documents>,  
[http://www.ataugaco.org/Sites/Autauga\\_County/Documents/Main/SKMBT-C36017071908210.pdf](http://www.ataugaco.org/Sites/Autauga_County/Documents/Main/SKMBT-C36017071908210.pdf)



requestors, we see significantly elevated burdens placed upon individuals who are attempting to engage their constitutionally protected right to inspect public court records in the form of criminal jury trial transcripts. In both our review of the current literature, our exemplification of present procedure, and our audit of the universe of policies, we find significant barriers to access of public records that functionally transform *de jure* public records into *de facto* non-public records. The consequences of these barriers function similarly to other forms of managerialized rights, where the process of vindicating and engaging those rights often serves to leave those rights functionally unprotected. This affects not only researchers looking to add knowledge in understudied disciplines and policymakers looking to evaluate current processes, but it is also an affront to the very notion of public access to public documents as protected both by common law understandings and the U.S. Constitution.

We have left aside for much of this paper the field of computational legal analysis and how the problem with retrieving criminal jury trial transcripts has anything to do with this rising field of scholarship, so we return to it here as a call for action. Computational legal analysis has a unique opportunity to make positive gains in the study of inequality and bias across court systems by using our unique methodological toolsets to make use of unconventional or unstructured data at scale. We argue that court transcripts are such a data frontier. In fact, we argue that criminal jury trial transcripts are currently the only formal records of many types of interacting discrimination that plague the criminal legal system. However, to contribute in these spaces, computational legal analysis cannot be satisfied to use what data is easily available. Instead, computational legal scholars must continue to interrogate the systematic barriers to data access and examine how those barriers may obfuscate inequity and sustain systems of inequality.

The current solution to the problem of public access to public court transcripts both individually and at scale is not an acceptable one. Presently, in order to obtain court transcripts at scale for research or as a member of the public who is legally entitled to access them, your most viable solution is winning tens of thousands, to millions of dollars from research foundations to pay the government for the transcripts that they are legally required to provide to you or to produce a networked connection to some sort of legal stakeholder to procure them for you. Instead, the government must assume the financial costs of making records public and the process-costs of making records not just *de jure* public, but also *de facto* public. This will require an overhaul of the current system, but a system that obfuscates legal processes and denies the public their legally protected right to inspect court data is not a system that deserves to remain.

### *1. Limitations*

This work is presently limited by looking at available policy information without testing the veracity of the policy statements. That is, here we presume that the policy as stated on various clerk of court websites is how the policy would be enacted in practice. We hope to lessen this limitation in our future work by actually requesting transcripts from a diverse sample of jurisdictions and recording how the process in reality matches or deviates with the policy we located. We hope that in addition to the substantive conclusions and data access gains from the transcripts themselves, we can make significant gains in the evaluatory research that is needed to fully analyze public data policies as they concern criminal jury trial transcripts.

Finally, this work is also necessarily interpreted through the lens of two legally trained coders. Both coders of this data have completed law school. This presents advantages in ensuring that all possibly relevant information was obtained for the audit, but almost certainly

outpaces the ability of the average layperson to retrieve the information cited here. In an effort to lessen the impact of our own expertise, we plan to conduct a coding test on a digital survey platform where laypeople will be tasked to retrieve information about the availability of court transcripts in different counties. Their results will then be compared to the results of the data audit described here.

#### IV. CONCLUSION

Court data is generally inaccessible to the public. Although records are legally public, the difficulty in obtaining those records makes them functionally not public at all. The managerialized rights framework elucidates how that works: the right to public records operates not as a guarantee from those who hold the records (the clerks, the courts) to make those records freely available but instead is held behind procedural barriers that functionally obstruct that right entirely. This is, of course, assuming that the records continue to exist for any meaningful length of time.

As discussed in this paper, that means that some areas of research are necessarily hindered, with important implications for full understanding of how the criminal legal system works. Limited access to data means limited access to understanding precisely how bias operates in the criminal legal system because we broadly lack empirical proof of patterns within the court system and tangible connection between those patterns and socio-demographic information.

We hold that clerks, as the office in charge of non-judicial aspects of the court, ought to take a more proactive public archival role in collecting and maintaining court records and data in order to take the responsibility for managing rights outside the hands of the people who have less perfect knowledge of the system. There are whispers of this type of movement happening

already, with various counties publicly taking steps to produce more transparent data or implementing more accessible portal-type tools to allow members of the public to easily request records. However, we urge that these public-facing tools be created with a more expansive understanding of what counts as public records and with consistency across jurisdictions. An ideal data portal would be responsive to the four themes identified in this audit: missing information, varied contact procedure, financial burdens and process inefficiency, and expert knowledge requirements on the part of requestors. Addressing these themes adequately is a substantial undertaking that requires courts to produce and maintain better data tools, be responsible for the cost of data production, and develop systems for efficient data dissemination. We argue that these changes are necessary to fulfill the duty of the courts under their own laws to make public data truly accessible to the public and available for the types of system-level analysis necessary to measure currently unmeasurable legal events and sources of legal system inequity.