16

The Civil Justice Data Gap

Tanina Rostain and Amy O'Hara

Every year, millions of people living on the margins cycle through a civil justice system that exacts an enormous toll on their housing security and financial well-being. Eviction rates reached crisis levels before the coronavirus pandemic and are expected to continue to climb as emergency moratoria have been lifted.[1] Debt-collection cases have not received the same national attention as evictions, but may represent a quarter to a third of state civil cases according to recent statistics.[2] The pandemic has contributed to a rise in consumer debt, creating a looming threat of increased collection cases.[3] Civil courts have become eviction and collection mills, whose apparent function is issuing judgments to help landlords and debt collectors grind down tenants and creditors.[4] According to available evidence, the civil justice

Thanks to James Carey, Amy Olivero, Garrett Quenneville, Anna L. Stone, Stephanie Straus, and Eoin Whitney for their contributions to this chapter and to the Georgetown Civil Justice Data Commons project. This work is supported by grants from the National Science Foundation (AWD 1952067) and The Pew Charitable Trusts.

[1] *See* Kathryn M. Leifheit et al., *Expiring Eviction Moratoriums and COVID-19 Incidence and Mortality*, 190 Am. J. Epidemiology 2503 (2021); Jacob Haas, Jasmine Rangel, Juan Pablo Garnham & Peter Hepburn, *Preliminary Analysis: Eviction Filing Trends after the CDC Moratorium Expiration*, Eviction Lab (Dec. 9, 2021), https://evictionlab.org/updates/research/eviction-filing-trends-after-cdc-moratorium/.

[2] There is a lack of comprehensive data of debt-collection suits across state court systems. As of 2013, an estimated 24 percent of all state civil cases were debt collection. Of the twelve states that reported court data about debt claims past 2013, nine reported that debt claims were the most common type of civil case. In Texas, the number of debt-collection suits more than doubled and constituted about 30 percent of Texas' civil docket between 2014 and 2018. *See* Pew Trs., How Debt Collectors Are Transforming the Business of State Courts (2020), https://www.pewtrusts.org/en/research-and-analysis/reports/2020/05/how-debt-collectors-are-transforming-the-business-of-state-courts.

[3] *See* Pamela Foohey, Dalié Jiménez & Christopher K. Odinet, *The Debt Collection Pandemic*, 11 Cal. L. Rev. Online 222 (2020).

[4] *See* Paula L. Hannaford-Agor et al., Nat'l Ctr. for State Cts., The Landscape of Civil Litigation in State Courts (2015), https://www.ncsc.org/__data/assets/pdf_file/0020/13376/civiljusticereport-2015.pdf.

system is failing in its promise to protect the rights of poor people against powerful private interests and entrenching them further in poverty.

But "available evidence" is sparse. We know little about how courts dispense justice or how court entanglements affect peoples' lives. Our knowledge of how civil justice institutions function, despite the system's societal centrality, is piecemeal and relies on extrapolation from site-specific or out-of-date studies. In a system where all parties are meant to be treated equally, we don't know whether the race, ethnic origin, or gender of litigants affects case outcomes. We can't determine whether there are biases in the system or where they are concentrated. We don't know what events precipitate civil justice involvement, nor the consequences that flow from it. We don't know how unrepresented individuals fare against businesses with representation. We don't even have an exact number of how many unrepresented litigants there are or in which types of cases they are concentrated[5] or whether the scarce resources that *are* available are being allocated effectively and equitably.

As a recent report by the American Academy of Arts and Sciences emphasizes, many institutions and individuals have an interest in answering questions like these about the functioning of the justice system.[6] State, county, and municipal courts are a central feature in the lives of low-income people and, increasingly, the middle class. Courts have an obligation to the people who pass through their doors to administer justice equitably. Litigants are entitled to know whether courts are living up to this ideal. Courts and legal services providers have particular interests in gaining operational insights about how well they serve the people before them and how effectively they allocate their resources. And the public, which funds the civil justice system, has a stake in knowing how well the civil justice system functions and whether its outcomes are just.[7] Erosion of public trust puts the legitimacy of this system at risk.[8]

In other public spheres, large datasets are producing new and actionable knowledge about the performance of public and private institutions. Health care and health insurance datasets have been linked to show that more years of childhood Medicaid eligibility correlate with fewer adult hospitalizations for Black Americans.[9] In Cuyahoga County, Ohio, the Center on Urban Poverty and Community

[5] *See SLRN Brief: How Many SRLs? (SRLN 2019)*, Self-Represented Litig. Network (Mar. 22, 2022), https://www.srln.org/node/548/srln-brief-how-many-srls-srln-2015 (estimating more than 75 percent of civil court cases have at least one self-represented litigant); *see also* Jessica Steinberg, *Demand Side Reform in the Poor People's Court*, 47 Conn. L. Rev. 741, 750–51 (2015) (in landlord-tenant matters typically 90 percent of tenants appear unrepresented while 90 percent of landlords appear with counsel).

[6] *See* Am. Acad. of Arts & Scis., Measuring Civil Justice for All: What Do We Know? What Do We Need to Know? How Can We Know It? (2021), https://www.amacad.org/publication/measuring-civil-justice-all.

[7] *See* Geoffrey McGovern & Michael D. Greenberg, Who Pays for Justice? (2014).

[8] *See* Am. Acad. of Arts & Scis., Measuring Civil Justice.

[9] Laura R. Wherry, Sarah Miller, Robert Kaestner & Bruce D. Meyer, *Childhood Medicaid Coverage and Later-Life Health Care Utilization*, 100 Rev. Econ. Stat. 287 (2018).

Development has used data from more than thirty-five different administrative programs to create two integrated data systems for research.[10] Linkages in these comprehensive data systems have shown increased shelter use by evicted families and identified downstream consequences such as childhood lead poisoning.[11]

As these projects show, big data are particularly powerful in the study of social institutions by delivering variety, volume, and velocity. Real-time data are captured for nearly all of our everyday whereabouts, purchases, searches, and interactions with private individuals and public institutions. Everyday interactions with other people and social institutions are chronicled by the terabyte through active or passive engagement with digital technologies. Data are created and available in near-real time.[12] These features produce broader knowledge than was possible with limited size and site-specific case studies, permitting both wide generalizations and highly granular views into the functioning of institutions and their effects on people's lives.

What explains the lack of data in civil justice? For one, not every court collects all case level information in electronic form. Some simply don't have the resources; others don't see value in doing so. Even where data exist, rules often restrict access. Courts are unwilling to share data, or uncomfortable doing so, for fear that vulnerable litigants will be harmed or that the data will be used to cast a critical eye on courts themselves.[13] Courts lack resources and incentives to improve their data, resulting in civil justice data that are incomplete, inaccurate, dirty, and disorganized. Despite the uniformity of needs among court administrators and a small handful of vendors providing case management systems, there are no widely used standards for data collection and harmonization. The absence of common standards makes it impossible to compare jurisdictions or even to study individual jurisdictions across different case types and subject matters.

The lack of high-quality, accessible data is a major deterrent to producing knowledge about civil justice. To obtain data, researchers must engage in a hunt-and-peck exercise that may or may not yield useful results. As a first step, they need to locate datasets that may be useful to them. Then, they determine whether there are legal limitations to obtaining the data. If the data are not publicly available but a

[10] *See CHILD Data System*, The Ctr. on Urb. Poverty & Cmty. Dev., https://case.edu/socialwork/povertycenter/data-systems/child-data-system; *see also Harnessing Data for Social Good*, NEOCANDO, https://neocando.case.edu/.

[11] Francisca García-Cobián Richter et al., The Cleveland Eviction Study: Downstream Paths of Evictions into Homelessness and Loss of Human Capital (2019), https://lasclev.org/wp-content/uploads/Costs-of-Eviction_11052019.pdf; Claudia Coulton et al., Downstream Consequences of Childhood Lead Poisoning: A Longitudinal Study of Cleveland Children from Birth to Early Adulthood (2020), https://case.edu/socialwork/povertycenter/sites/case.edu.povertycenter/files/2020-07/Downstream_06182020_rev07082020.pdf.

[12] Rebecca A. Johnson & Tanina Rostain, *Tool for Surveillance or Spotlight on Inequality? Big Data and the Law*, 16 Ann. Rev. L. & Soc. Sci. 453 (2020).

[13] *See infra* Section 16.2.

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

court has discretion to share them, researchers must engage in protracted negotiations to obtain access. Once researchers have access, they must spend weeks cleaning the data. The process of finding and obtaining data and making them usable deters researchers from studying the civil justice system and impedes the development of evidence-based policies in the civil justice arena. Although data science is producing new insights about many social institutions, few data scientists are working in the civil justice field.

This chapter considers how improving data collection, sharing data, and ease of data analysis can make civil justice institutions more accountable to government institutions, their constituents, and the public at large, and create opportunities for civil justice reform.

Section 16.1 starts by identifying key questions that analysts are well positioned to address. These include problems of structural racism and bias against people of color, the advantages of being represented, and the antecedents and consequences of entanglement in the civil justice system. Section 16.2 considers the obstacles to increasing knowledge about civil justice problems. These obstacles include the lack of good (or any) data, legal barriers to obtaining data, and real and perceived institutional risks to sharing data.

Section 16.3 then describes a civil justice data commons as one promising avenue that the authors are prototyping to address current access-to-justice challenges[14] and free up the empirical opportunities offered by civil justice data.[15] This Civil Justice Data Commons (CJDC) functions as a trusted intermediary among data sharers and data users. The CJDC's technical infrastructure is automated to satisfy conditions for the supply of data and the demands of users. To facilitate access for researchers – speeding up the evidence-building process – the CJDC hosts documented, harmonized data, cleans and links files as needed, and provides tools and analytic software for research. The data commons allows for faster access to data for academic and policy research, as well as operational insights for courts and civil justice institutions to improve equity and service.

Several different tech trends examined in this volume – from ODR to online courts to e-filing to the growing menu of legal tech tools – will transform civil justice in the years to come. Each will increase both the amount of and focus on data. A key

[14] *See* Robert L. Grossman et al., *A Case for Data Commons: Toward Data Science as a Service*, 18 Computing Sci. Eng'g 10 (2016); *see also* Robert L. Grossman et al., *The Design of a Community Science Cloud: The Open Science Data Cloud Perspective*, 2012 High Performance Computing, Networking, Storage & Analysis 1051 (2012); Margaret Hagan, Jameson Dempsey & Jorge Gabriel Jiménez, *A Data Commons for Law*, Legal Design & Innovation (Apr. 1, 2019), https://medium.com/legal-design-and-innovation/a-data-commons-for-law-60e4c4ad9340.

[15] *See* Robert L. Grossman, *How Data Commons Can Support Open Science*, Sage Bionetworks (2019), https://sagebionetworks.org/in-the-news/how-data-commons-can-support-open-science/; *see also* Charles Q. Choi, *Migrating Big Astronomy Data to the Cloud*, 584 Nature 159 (2020).

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

question is whether the civil justice system – and the judges, lawyers, litigants, and justice partners who inhabit it – can harness data to promote the just, equitable, and efficient administration of justice. A civil justice data commons can help answer that question.

16.1 IMPORTANT QUESTIONS ABOUT THE FAIRNESS OF THE CIVIL JUSTICE SYSTEM

16.1.1 *Structural Racism and Bias*

In the criminal justice context, abundant high-quality policing and court data have allowed researchers to document the scope and depth of structural racism and other biases. This research underscores what smaller-scale observational and quantitative studies had previously shown: Black Americans are systematically treated worse than white Americans at every juncture of the criminal justice system. Policing data have shown that Black people are more likely to be subjected to roadside stops and are more likely to be the victims of lethal force.[16] Court datasets have allowed researchers to show that Black drivers are likely to pay higher traffic fines than white drivers[17] and that judicial mechanics and the peremptory challenge system adversely affect Black defendants.[18] If you are Black, you are more likely to receive a longer sentence than a white person in a plea deal.[19] These studies confirm the systemic pervasiveness of racism in criminal justice. Data access permits studies with sufficient granularity to pinpoint specific areas where race might have the strongest magnifying effect, which may suggest priorities for formulating and testing reforms.

Large-scale studies that investigate the impact of race, ethnic origin, and gender in the civil justice system are sparse but show promise.[20] One major foray comes from the Princeton Eviction Lab, which published the first study documenting the

[16] Emma Pierson et al., *A Large-Scale Analysis of Racial Disparities in Police Stops across the United States*, 4 NATURE HUM. BEH. 736 (2020); James W. Buehler, *Racial/Ethnic Disparities in the Use of Lethal Force by US Police, 2010–2014*, 107 AM. J. PUB. HEALTH 295 (2017); Cody T. Ross, *A Multi-Level Bayesian Analysis of Racial Bias in Police Shootings at the County-Level in the United States, 2011–2014*, 10 PLOS ONE e0141854 (2015).

[17] Felipe Goncalves & Steven Mello, *A Few Bad Apples? Racial Bias in Policing*, 111 AM. ECON. REV. 1406 (2020).

[18] Scott Kostyshak & Neel Sukhatme, *Down to the Last Strike: The Effect of the Jury Lottery on Criminal Convictions* (Geo. L. Fac. Publ'ns & Other Works, Working Paper No. 2156, 2020).

[19] *See* Angela J. Davis, *Racial Fairness in the Criminal Justice System: The Role of the Prosecutor*, 39 COLUM. HUM. RTS. L. REV. 202, 205–11 (2007).

[20] *See* Matthew Desmond, *Eviction and the Reproduction of Urban Poverty*, 118 AM. J. SOCIO. 88 (2012); Peter Hepburn, Renee Louis & Matthew Desmond, *Racial and Gender Disparities among Evicted Americans*, 7 SOCIO. SCI. 649 (2020); BRIAN J. MCCABE & EVA ROSEN, EVICTION IN WASHINGTON, DC: RACIAL AND GEOGRAPHIC DISPARITIES IN HOUSING INSTABILITY (2020), https://www.streetsensemedia.org/wp-content/uploads/2020/10/Georgetown-University-DC-Eviction-Report-2014-2018.pdf.

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

demographics of evicted renters on a national scale. Using a dataset containing 4.1 million records from public, private, and commercial sources, the study showed Black renters experienced the highest rate of eviction filings and judgments, and Black and Latin female renters faced higher eviction rates than their male counterparts.[21] Data-driven studies like those by the Eviction Lab can demonstrate the magnitude of inequity in the civil justice system. Their study is a significant step in understanding which demographic groups in the United States are suffering the brunt of the eviction crisis and raises new questions about how the civil legal system plays into this crisis. Do differing eviction rates among racial and gender groups mirror structural biases in society at large? Or does the civil justice system magnify these problems? Accessible civil justice data are needed not only to understand the system's role in replicating or worsening structural biases in eviction cases, but also its role in consumer debt and other areas that disproportionately affect poor people, people of color, and women. Population-level datasets contain large numbers of observations and variables to help pinpoint court contacts and processes that have the greatest impact.

16.1.2 *When Do Lawyers Make a Difference (and Why?)*

Civil justice data can also shed light on whether lawyers make a difference in outcomes and, if so, in what types of cases and why.[22] To date, studies comparing represented and non-represented parties have been site-specific and relied on observational data or randomized controlled trials. Consequently, their findings are not easily generalized to other subject matters and other settings. Large-scale court datasets will allow comparisons across courts, subject matter areas, types of representation, and jurisdictions, producing more nuanced knowledge with broader impact.

The few studies investigating whether being represented by a lawyer affects the outcome of a case tend to show that representation makes a difference, but how much difference it makes depends in large part on the type of service provided, case type, and procedural complexity of the case.[23] Some research suggests that, in select

[21] Hepburn, Louis & Desmond, *Racial and Gender Disparities among Evicted Americans*.

[22] *See generally* Rebecca Sandefur, *The Impact of Counsel: An Analysis of Empirical Evidence*, 9 Seattle J. for Soc. J. 51 (2010).

[23] *See* Ellen Degnan, Thomas Ferriss, D. James Greiner & Roseanna Sommers, Trapped in Marriage (Oct. 23, 2018) (unpublished manuscript), https://papers.ssrn.com/sol3/papers.cfm?abstract_idwww3277900 (suggesting that legal representation may have varied results based on setting and form of service offered); D. James Greiner, Cassandra W. Pattanayak & Jonathan P. Hennessy, *The Limits of Unbundled Legal Assistance: A Randomized Study in a Massachusetts District Court and Prospects for the Future*, 126 Harv. L. Rev. 901 (2013) (finding full representation for eviction made a difference as compared to limited assistance); D. James Greiner & Cassandra Wolos Pattanayak, *Randomized Evaluation in Legal Assistance: What Difference Does Representation (Offer and Actual Use) Make?* 121 Yale L.J. 2118 (2012) (reporting that an offer of full representation by a law school clinic in unemployment benefits

settings, being represented by a non-lawyer advocate confers similar advantages to being represented by a lawyer. These studies raise important questions about how courts should deal with cases where one or both parties are not represented by a lawyer and how legal services providers should prioritize the provision of their resources.

Data-driven research confirming and extending these findings offers insights into how to "level the playing field" between represented and unrepresented parties. Some jurisdictions have changed their rules to allow judges a more active role in explaining the law or eliciting facts to a pro se party.[24] Courts might also focus on simplifying procedures or prohibiting lawyers in certain matters or fora altogether.[25]

Research showing whether and when being represented affects case outcomes can also be important for legal services providers allocating resources. In legal aid settings, eligible applicants may receive representation depending on the availability of a lawyer specializing in their type of case. Research on the differential effects of representation might persuade legal services providers, which turn away almost half of eligible applicants,[26] to shift their resource priorities. Some providers report that they have already begun targeting their resources to clients with income levels just below eligibility cut-offs after data-based research showed representation did not affect outcomes at lower income levels.[27] Large-scale data studies about the impact of lawyers would also provide evidence for or against various approaches being promoted to increase access to the civil justice system. With the eviction crisis

eligibility determination appeals delayed proceedings by two weeks; impact of actual use was unclear); Rebecca L. Sandefur, *Elements of Professional Expertise: Understanding Relational and Substantive Expertise through Lawyers' Impact*, 80 Am. Socio. Rev. 909 (2015); Sandefur, *The Impact of Counsel*, at 73 (across forty-five studies, lawyers have significant impact on case outcome, but effect did not correlate with legal expertise; greatest effect was in procedurally complex cases); Anna E. Carpenter, Alyx Mark & Colleen F. Shanahan, *Trial and Error: Lawyers and Nonlawyer Advocates*, 42 Law & Soc. Inquiry 1023 (2017) (legal representation has benefits in unemployment cases, but non-lawyer advocates are less equipped to challenge judges or advance novel claims).

[24] *See Self-Representation on Justice Index*, Nat'l Ctr. for Access to Justice, https://ncaj.org/state-rankings/2020/self-representation.

[25] *See* Steinberg, *Demand Side Reform*, at 786–802; *see also An Overview of Small Claims Court*, Mich. Legal Help, https://michiganlegalhelp.org/self-help-tools/money-and-debt/overview-of-small-claims-court (stating that no lawyers are allowed in small claims court in Michigan); *Basic Considerations and Questions*, State of Cal. Dep't of Consumer Affs., https://www.dca.ca.gov/publications/small_claims/basic_info.shtml (in California small claims court, rules are simplified, hearings are informal, and lawyers are typically not permitted).

[26] Legal Servs. Corp., The Justice Gap: Measuring the Unmet Civil Legal Needs of Low-Income Americans 42–44 (2017), https://www.lsc.gov/justicegap2017.

[27] Interview with Jim Sandman (Jan. 27, 2021); *see, e.g.*, Rachel Dissell, *Legal Aid "Securing Stability" for Low-Income Clients, Report Shows*, Cleveland.com (June 16, 2019),https://www.cleveland.com/news/2019/06/legal-aid-securing-stability-for-low-income-clients-report-shows.html (Cleveland's Legal Aid turns away about half of qualifying low-income clients due to lack of resources).

worsening, the movement to fund lawyers to represent tenants at risk of eviction is gaining ground.[28]

Such studies might also be helpful in answering the question of which types of services make a difference. In the last ten years, limited scope representation has been promoted as an important means of increasing access to justice by allowing fewer lawyers to represent more people, an approach that many legal services providers are pursuing on the assumption that they can help more clients.[29] Despite the growth of brief advice services,[30] there is little evidence that these services have a positive effect on outcomes. One study found that while brief advice assisted clients in surmounting procedural barriers, it did not have significant effects on substantive outcomes.[31] However, this study, which focused on the evictions caseload of a legal aid provider in San Mateo County during the summer of 2009, may not be generalizable to other settings.

Improving short-term outcomes – helping a client win a specific case, stay in their home, or avoid wage garnishment – is one measure of effective representation. Others include helping clients understand the law, supporting them through the process, and empowering them to advocate for themselves.[32] An understanding of whether and when representation makes a difference is important knowledge to have in allocating scarce resources among eligible applicants who live in poverty and seek legal assistance.

16.1.3 *Civil Justice Involvement and Poverty*

The questions we have identified so far focus on whether civil justice institutions treat people fairly, allocate resources equitably, and achieve just results. Other research avenues focus on the people involved, not the court encounters themselves.

[28] Seven cities have enacted laws guaranteeing the right to counsel for low-income people in eviction cases: Cleveland, OH; New York City, NY; Boulder, CO; San Francisco, CA; Baltimore, MD; Newark, NJ; and Philadelphia, PA. *See* Sandra Park & John Pollock, *Tenants' Right to Counsel Is Critical to Fight Mass Evictions and Advance Race Equity during the Pandemic and Beyond*, ACLU News & Comment. (Jan. 12, 2021), https://www.aclu.org/news/racial-justice/tenants-right-to-counsel-is-critical-to-fight-mass-evictions-and-advance-race-equity-during-the-pandemic-and-beyond; *see also* Beth Harrison, *District of Columbia Joins Right to Counsel Movement with New Eviction Defense Funding*, Mgmt. Info. Exch. J. (Fall 2017), at 41, 41–42.

[29] *See* Jessica K. Steinberg, *In Pursuit of Justice? Case Outcomes and the Delivery of Unbundled Legal Services*, 18 Geo. J. Poverty Law & Pol'y 453, 453–56 (2011).

[30] *See* Legal Servs. Corp., The Justice Gap, at 31–32 (citing that 40 percent of low-income people seeking professional legal help receive only legal advice; only 20 percent receive full representation in court); *see also* NPC Rsch., Sargent Shriver Civil Counsel Act Evaluation 3 (2020), https://www.srln.org/node/1515/report-sargent-shriver-civil-counsel-act-report-legislature-judicial-council-california (stating that between 2012 to 2020, 39 percent of low-income litigants seeking civil legal help in California received "limited-scope legal assistance").

[31] *See* Steinberg, *In Pursuit of Justice?*

[32] *See* Catherine R. Albiston & Rebecca L. Sandefur, *Expanding the Empirical Study of Access to Justice*, 2013 Wis. L. Rev. 101 (2013).

Getting a Fuller Picture from Big Data. More accurate and consistent data provide opportunities for deeper understanding about individuals' interactions with the civil justice system: who enters the system (and why), how they navigate the system, and what impacts this interaction has on them in the long term. Specifically, linked population-level datasets that consider the antecedents and consequences of court involvement can widen the lens of research, leading to inferences and interventions that can scale across sites. By linking information from before and after court involvement, it may be possible to understand mechanisms that help avoid entanglement in the justice system or to mitigate its negative effects. This research shifts the emphasis away from court processes and legal resources to understanding and addressing the conditions and events that lead up to court involvement to begin with. In doing so, it may help people avoid being evicted by landlords or sued by debt collectors altogether.[33]

This type of analysis is already being done at other intersections of poverty and policy. For example, researchers in Sonoma County, California, integrated four years of county and state records from health, mental health, substance abuse, housing, criminal justice, and human services systems.[34] This cross-domain dataset showed that 1 percent of the population accounted for one-quarter of jail time and behavioral health services and half of overnight homeless housing.[35] Another study linked administrative records from the Child Care and Development Fund, a federal subsidy program, with American Community Survey data to reveal that women who worked full-time but had less social support were more likely to receive the child care subsidy.[36] The study considered characteristics of the mother, such as education level, attainment of labor skills, and community ties, and found that relative to other low-income women, the women who received the child care subsidy tended to have higher educational attainment (high school diploma or higher).[37]

Understanding the long-term consequences of a government intervention has also been made possible with linked datasets. One recent study created more than 150 million parent-child links to analyze the downstream intergenerational effects

[33] *See* Colleen F. Shanahan & Anna E. Carpenter, *Simplified Courts Can't Solve Inequality*, 148 Daedalus 128 (2019) (describing how court simplification reforms mitigate but do not solve the underlying problems of inequality and the failures of the legislative and executive branches to provide a social safety net).

[34] Elsa Augustine & Evan White, Cal. Pol'y Lab, High Utilizers of Multiple Systems in Sonoma County (2020), https://www.capolicylab.org/wp-content/uploads/2020/07/High-Utilizers-of-Multiple-Systems-in-Sonoma-County.pdf.

[35] Specifically, the 1 percent of the population accounted for 25 percent of jail time, 28 percent of annual behavioral health costs, 52 percent of overnight homeless housing. *Id.* at 6.

[36] Rachel Shattuck, *High Labor Force Attachment, but Few Social Ties? Life-Course Predictors of Women's Receipt of Childcare Subsidies*, U.S. Census Bureau Ctr. for Econ. Stud., Working Paper No. CES-19-26, 2019, https://www.census.gov/library/working-papers/2019/adrm/ces-wp-19-26.html.

[37] *Id.*

of the 1970 Clean Air Act Amendment and found that regulation-induced air quality improvements during pregnancy correlated with parental investment and higher college attendance in later generations.[38] This analysis was achieved through linking decennial census data, administrative records, and survey data and utilizing Protected Identification Keys (PIKs) in order to anonymize personal identifiers.[39] In another study, researchers linked national administrative data on the labor market and earnings with records on public housing to find that children who resided in public housing that was demolished later earned more than children who remained in non-demolished public housing, suggesting that job accessibility improved for children forced to relocate.[40]

The Interaction between Poverty and Entanglement with the Justice System. Just as these studies used linked data to understand the role of government policies in the lives of people living in poverty, research linking court datasets to datasets that are indicators of health, financial circumstances, and other measures of well-being can yield knowledge about the effect of court entanglement over the course of a person's life. Research is already under way. The Institute for Research on Poverty at the University of Wisconsin–Madison found women with an incarcerated family member have a higher risk of poor health, and Black women are, in turn, disproportionately likely to have an incarcerated family member.[41] Researchers looking at Connecticut evictions found persistent, largely untreated, mental health problems and continuing housing instability.[42] Other research, using mixed methods and blending administrative data with court data, shows correlations between eviction and criminal justice involvement,[43] how neighborhood disadvantage can exacerbate racial health disparities,[44] and how eviction in early childhood is associated with neighborhood poverty, food insecurity, and obesity in later childhood and adolescence.[45]

[38] Jonathan Colmer & John Voorheis, *The Grandkids Aren't Alright: The Intergenerational Effects of Prenatal Pollution Exposure*, U.S. Census Bureau Ctr. for Econ. Stud., Working Paper No. CES-20-36, 2020, https://www.census.gov/library/working-papers/2020/adrm/CES-WP-20-36.html.

[39] *Id.* at 11.

[40] John C. Haltiwanger et al., *The Children of HOPE VI Demolitions: National Evidence on Labor Market Outcomes*, U.S. Census Bureau Ctr. for Econ. Stud., Working Paper No. CES-20-39, 2020, https://www.census.gov/library/working-papers/2020/adrm/CES-WP-20-39.html.

[41] Hedwig Lee et al., *The Effects of Having an Incarcerated Family Member on Black Women's Health*, 36 Focus 22, 22–29 (2020).

[42] Jack Tsai, Natalie Jones, Dorota Szymkowiak & Robert A. Rosenheck, *Longitudinal Study of the Housing and Mental Health Outcomes of Tenants Appearing in Eviction Court*, 56 Soc. Psychiatry & Psychiatric Epidemiology 1679 (2020).

[43] Aaron Gottlieb & Jessica W. Moose, *The Effect of Eviction on Maternal Criminal Justice Involvement*, 4 Socius 1 (2018).

[44] Douglas S. Massey et al., *Neighborhood Disadvantage and Telomere Length: Results from the Fragile Families Study*, Russell Sage Found. J. Soc. Scis., Apr. 2018, at 28.

[45] Kathryn M. Leifheit et al., *Eviction in Early Childhood and Neighborhood Poverty, Food Security, and Obesity in Later Childhood and Adolescence: Evidence from a Longitudinal Birth Cohort*, SSM – Population Health, August 2020, at 100575.

Studies that investigate correlations between involvement with the civil justice system and other public institutions are beginning to emerge. For example, a study linking bankruptcy filings to administrative tax and foreclosure data found that filers who received Chapter 13 bankruptcy protections experienced increased annual earnings and decreased foreclosure rates.[46] Several studies have also appeared that focus on the short- and long-term consequences of being evicted for poor families. Using New York City housing court records, researchers found that evictions correlated with an increased risk of homelessness, long-term residential instability, and emergency room use but did not significantly worsen longer-term financial circumstances.[47] A longitudinal comparison of Cook County court records to credit bureau and payday loans data illustrated how the effects of successful evictions were small relative to the financial strain experienced by both evicted and non-evicted tenants in the time leading up to eviction actions.[48] Taken together, these studies raise the possibility that the prevailing view (which assumes that preventing eviction will have long-term benefits on alleviating poverty) may be myopic and that the locus of policy should shift from court processes to poverty-reduction programs.

More and greatly accelerated research using large, linked datasets is needed to determine the consequences of court involvement for the life cycle of poverty across sites and population subgroups. Providing lawyers and legal assistance to poor litigants assumes that the best locus of intervention to alleviate the effects of poverty must be the civil justice system – and that more lawyers and more legal assistance is the best approach. More studies may tell us, however, that assistance and opportunities upstream may be the most effective way of addressing housing and financial insecurity to prevent eviction and consumer debt actions from being initiated in the first place. Studies may also show that the impact of representation varies across race, ethnic enclaves, age groups, and population density.

Easily accessible civil justice datasets are needed to undertake such studies. In Section 16.2, we describe current challenges to obtaining such data and explain how the development of a civil justice data commons may begin to overcome these barriers.

[46] Will Dobbie, Paul Goldsmith-Pinkham & Crystal S. Yang, *Consumer Bankruptcy and Financial Health*, 99 R. Econ. & Stat. 853 (2017).

[47] Robert Collinson & Davin Reed, The Effects of Evictions on Low-Income Households (Dec. 2018) (unpublished manuscript), https://www.law.nyu.edu/sites/default/files/upload_documents/evictions_collinson_reed.pdf (finding some evidence evictions modestly decrease earnings and little evidence evictions increase receipt of public assistance or significantly worsen employment comes, thus suggesting eviction prevention policies could provide consumption smoothing benefits for low-income households but are unlikely to significantly reduce poverty in isolation).

[48] John Eric Humphries, Nicholas S. Mader, Daniel I. Tannenbaum & Winnie L. van Dijk, *Does Eviction Cause Poverty? Quasi-Experimental Evidence from Cook County, IL*, (Nat'l Bureau of Econ. Rsch., Working Paper No. 26139, 2019), https://doi.org/10.3386/w26139.

16.2 BARRIERS TO ACCESSING DATA

An overarching challenge with civil justice data is the disaggregation of authority over civil dockets among states, counties, and municipal jurisdictions. In some states, managerial authority over the court system, and therefore control over data collection and sharing practices, is centralized in the supreme court or chief court administrative office. This makes it possible for the court to dictate that the same data be collected under the same standards throughout the state's civil courts. As former Chief Justice Bridget McCormack notes in her chapter in this volume, Michigan has adopted a uniform system that is spreading to all its counties. Connecticut has also implemented a similar approach.[49] Other states have decentralized systems where each court has the power to dictate what data are collected. In some states, courts in different jurisdictions throughout the state have the option of maintaining their official records in either paper or electronic form.[50] This results in data that are siloed in multiple institutions and starkly unstandardized. Accessibility is governed by a huge and bewildering hodgepodge of laws and policies, rendering their availability for research limited. These barriers must be addressed if data-driven approaches are to produce knowledge about civil justice institutions.[51] This section describes five of the largest barriers to access.

16.2.1 *Data Disarray*

Although courts are moving to electronic case management systems, many are still managing case files the old-fashioned way: paper. Others are using case management systems, but case pleadings are handwritten hard copies scanned into the digital system and are not machine-readable.

The absence of machine-readable pleadings, which contain valuable information about the case and the litigants, is a major technical impediment to granular research on the civil justice system. To overcome this barrier, some researchers have resorted to hand-coding pleadings, an expensive and time-consuming process.[52]

[49] *See* Chapter 13 in this volume; *About Connecticut Courts: Administration and Operation of the Courts*, STATE OF CONN. JUD. BRANCH, https://www.jud.ct.gov/ystday/adminop.html.

[50] *See Access to Electronic Court Records*, CAL. CTS.: THE JUD. BRANCH OF CAL., https://www.courts.ca.gov/42512.htm.

[51] *See, e.g.*, Robert M. Goerge, *Data for the Public Good: Challenges and Barriers in the Context of Cities*, *in* PRIVACY, BIG DATA, AND THE PUBLIC GOOD: FRAMEWORKS FOR ENGAGEMENT 153 (Julia Lane et al. eds., 2014) (discussing the opportunity for states to be data stewards, responsible for how data assets are used and transformed into public goods. Administrative data from government agencies in health, education, social services, criminal justice, and employment could be analyzed by social scientists and policy actors to improve delivery of services).

[52] MCCABE & ROSEN, EVICTION IN WASHINGTON, DC, at 35 (explaining how McCabe and Rosen hand-coded PDFs because the other records didn't include "key pieces of information,

Over time, the adoption of e-filing that is integrated with case management systems may alleviate this problem (though many mandatory e-filing systems exempt unrepresented parties).[53] For now, it is necessary to hand-transcribe the information or develop optical character-recognition tools that can translate pleadings that are marked with stamps and often contain handwriting into machine-readable form.

Administrative case level data scraped from online court case search websites – which pull from case management system data originally entered by hand by a court clerk – can be messy and contain incorrect or inconsistent values, making them difficult to interpret. Addressing these issues often requires a manual comparison with the corresponding docket by a specialist who can align concepts properly. That is tedious when the dockets are available online and onerous when it is necessary to contact the clerk of the court to understand which dispositions numeric codes refer to. Improving data capture is critical to improving analyses and the function of courts.

16.2.2 *Barriers to a Common Taxonomy*

Another data challenge is the absence of a common taxonomy among courts. The lack of shared substantive and procedural classification systems results from the diverse sites of authority among state, county, and lower-level courts. Different definitions across courts, even for concepts as fundamental as indigency, hamper coding and comparisons. Court procedures with different terms for various pleadings, requirements, and judgments also complicate matters. A court may lump evictions and contract cases together under "contracts," or it may separate evictions and contracts but not distinguish among consumer debt cases (with some number of these classified separately under "small claims"). A rare few might use more granular categories that distinguish medical debt, consumer debt, and other debts, as an example. Coding these types of categories is necessary to study correlations between court involvement and life-course events. For instance, being able to identify medical debt is necessary to understand how much healthcare debt collection has grown since the pandemic.

In an attempt to tame some of this chaos, the National Center for State Courts (NCSC) developed the National Open Data System (NODS), intended to guide data collection in courts throughout the United States.[54] NCSC's exhaustive effort yielded standards for basic elements such as case type, representation by counsel,

including legal representation for landlords, subsidy status of tenants, or the amount of rent owed by the tenant").

[53] *See, e.g.*, *E-filing – Public*, Md. Cts., https://www.courts.state.md.us/mdec/efilingpublic ("E-filing is *not* mandatory if you do not have a lawyer."); Cal. R. Ct. 8.71 (self-represented litigants are *not required* to file documents electronically but may choose to do so).

[54] *National Open Court Data Standards*, Nat'l Ctr. for State Cts., https://www.ncsc.org/services-and-experts/areas-of-expertise/court-statistics/national-open-court-data-standards-nods.

case filing, disposition, and case closure date. NODS is an ambitious effort to transform the collection and classification of court data. Its adoption will be dictated by the cost of implementing or reconfiguring court case management systems and the extent to which courts believe that its implementation is consistent with the court's priorities.

While NODs is a forward-looking strategy, there may be other strategies to rationalize court records from different jurisdictions that have already been collected. One approach might be to use natural language processing to identify subcategories in court documents to generate classifications. These generated classifications could then be matched to NODS categories. For instance, an algorithm might be trained to recognize medical debt from other types of debt based on the language in pleadings. Whether such an approach could capture subcategories of case types or definitional and procedural variation – likely more challenging – remains to be seen.

16.2.3 *The Special Challenge of Assigning Demographic Characteristics*

A particular challenge involves obtaining demographic information for civil court participants – a prerequisite for understanding equity concerns within the civil justice system. Under NODS, civil courts are supposed to collect race, ethnicity, and other demographic data about litigants. Most courts do not currently capture these data, but even if they attempted to, they would face a fundamental difficulty: Many, if not most, defendants in certain types of civil actions do not appear in court. One recent study of debt-collection cases in state courts found that, for courts where there was available data, 70 percent of debt cases ended in default judgments.[55] With such high rates of default, it is impossible to collect information about the race, ethnic origin, or gender of the tenant or debtor using only court data.

One promising strategy to append demographic characteristics to court data is through data linkage to census or commercial data. Linking at the per-person level requires the court records to have complete, accurate name data or other unique identifiers to make a match. Previous linkages to census data have had match rates between 40 and 90 percent, depending on the completeness of the names and addresses available, with the higher match rates possible when additional fields such as age or date of birth are available. The sparse identifiers present on civil court records have led other researchers to incorporate additional data sources. For example, Case Western Reserve University's Center on Urban Poverty and Community Development matched Cleveland Housing Court data to Ohio

[55] *See* Pew Trs., How Debt Collectors Are Transforming the Business of State Courts (2020), https://www.pewtrusts.org/en/research-and-analysis/reports/2020/05/how-debt-collectors-are-transforming-the-business-of-state-courts.

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

Public Assistance records to obtain race information.[56] While this linkage allowed researchers to add otherwise absent demographic characteristics to eviction files, they were able to do so only for defendants who applied for and obtained public assistance.

16.2.4 *Legal Barriers*

While individual case files are available either online or in the clerk's office in every jurisdiction, legal rules may limit access to bulk electronic case records. Although all fifty states and the District of Columbia have enacted public records laws, more than a third exempt the judiciary from coverage.[57] Even when public record laws apply to the judiciary, it is unclear whether they entitle people to obtain electronic case files in bulk form.

Approximately thirty-two states have rules governing the availability in bulk form of electronic public case records. Some states like Arizona make case file records generally available for bulk download.[58] Even in states that do not bar access to case data, county courts may have rules that limit their accessibility.[59] Some states prohibit the dissemination of bulk court records in electronic form, except where explicitly provided by a court rule or order.[60] At least one state disseminates bulk court data only to commercial purchasers. Other states, in contrast, bar the bulk download of court files for commercial gain.[61] Finally, there are states like Hawaii, which will grant some requests for bulk data as long as the entity making the request is willing to pay for fees associated with the cost of providing the data.[62] Many states also charge a fee or have terms of service that prohibit the scraping of court websites or CAPTCHA intended to prevent it.[63] On top of this confusion,

[56] García-Cobián Richter et al., The Cleveland Eviction Study.

[57] *See State Freedom of Information Laws*, Nat'l Freedom of Info. Coal., https://www.nfoic.org/coalitions/state-foi-resources/state-freedom-of-information-laws.

[58] Ariz. Code of Jud. Admin. § 1-605; Ariz. R. Sup. Ct. 123.

[59] *See, e.g., Charleston County Public Index*, Charleston Cnty, S.C., https://jcmsweb.charlestoncounty.org/publicindex. For example, in Charleston and Greenville County, South Carolina, in order to access the court records, you have to accept a disclaimer that reads in part: "Access to the South Carolina Judicial Department Public Index web sites by a site data scraper or any similar software intended to discover and extract data from a website through automated, repetitive querying for the purpose of collecting such data is expressly prohibited." *Id.*

[60] *See, e.g.*, Elec. Access Pol'y for Circuit Ct. Recs. of the Ill. Cts. § 4.40.

[61] *See* Fla. Admin. Order No. AOSC14–19 (amended May 23, 2014); Mo. Ct. Operating Rule 2, Pub. Access to Record of the Jud. Dep't 2.10.

[62] *See* Haw. Ct. Recs. Rules 2.5, 10.16, 10.17.

[63] *See Just One Look: Alabama's ON-DEMAND Public Access to Trial Court Records*, Ala. Cts., https://pa.alacourt.com/default.aspx (illustrating that the state of Alabama charges $9.99 for a name search that includes one case detail; $9.99 for a case number search that includes one case detail; $5.00 for the first twenty pages of images and $0.50 per page thereafter; and $19.99 for case monitoring for the lifetime of the case if it is a district case and $29.99 if it is a circuit

states may have different privacy policies for court records with varying levels of restrictions.[64]

16.2.5 *Institutional Barriers*

Institutional barriers are arguably the most important to address to facilitate data sharing. In preliminary planning interviews for the CJDC, we found a wide spectrum of views among judges and court administrators about the value of collecting and sharing data. Some judges did not view the data as valuable for either research or operational purposes; others strongly believed that research on aspects of the civil justice system was important for determining how well the courts served their communities.

Courts also fear scope creep resulting in data misuse, where a project approved to study one topic veers to another and publishes a study that the data may not support. Safeguards and reviews of approved uses must keep pace with growing volumes of data. Courts may worry that errors in their data will come to light. The errors may result from hurried or flawed collection methods, or through poor data management, statistical analysis, logic, or communication. When errors arise, it is important to examine their origins and impacts.[65]

Even when court rules provide for public access to data, court personnel face administrative burdens to sharing them. They receive numerous requests for data and are pressed to develop internal custom-made approval processes on the fly. Most courts lack a data governance committee. Providing requested data is a time-consuming and mostly manual process fraught with risks of data misuse, errors in court data, and embarrassing findings.

16.3 WHAT A CIVIL JUSTICE DATA COMMONS OFFERS

A data commons provides institutions and policy makers a one-stop shop for both high-level and detailed information about civil justice problems, by giving institutions simple, secure methods to share their data with researchers, policy makers, and other civil justice stakeholders and researchers fast, frictionless, and facilitated

case); *New User Registration*, HARRIS CNTY. DIST. CLERK, https://www.hcdistrictclerk.com/eDocs/Secure/Registration.aspx; *Philadelphia Municipal Court Electronic Filing System*, PHILA. CTS. FIRST JUD. DIST. OF PA., https://fjdclaims.phila.gov/phmuni/login.do; *About This Site*, WASH. CTS., https://dw.courts.wa.gov/index.cfm; *Online Court Records Search*, BAKER CNTY. CLERK OF CT., www.civitekflorida.com/ocrs/app/search.xhtml; *Public Access to Court Information*, ARIZ. JUD. BRANCH, https://apps.supremecourt.az.gov/publicaccess/(X(1)S(ncwwtm45cm24di25fm5tozvl))/caselookup.aspx; *eAccess Terms and Conditions*, ARIZ. JUD. BRANCH, https://azcourtdocs.gov/arizona/publicTerms.admin.

[64] *See, e.g.*, State of Me. Supreme Jud. Ct., Me. Rules of Elec. Ct. Sys. Rule 1(B)(2).

[65] *See* Andrew W. Brown, Kathryn A. Kaiser & David B. Allison, *Issues with Data and Analyses: Errors, Underlying Themes, and Potential Solutions*, 115 PROC. NAT'L ACAD. SCIS. 2563 (2018).

data access. By centralizing data sharing and research functions, the commons can serve as a meeting place for analysts, and a space where participants can share subject matter expertise, technical tools, and research methods.

16.3.1 *What Is the CJDC?*

The civil justice data commons (CJDC) is built according to principles that underlie data commons in the health sciences, created to accelerate research on medicine and disease using large databases.[66] Like genomic and other health data commons, the CJDC is controlled by a governance regime negotiated between data sharers, data users, and the directors of the commons.[67] The terms of service reflect the data security needs and project scoping requirements of data contributors with access demands of researchers.[68] The governance model is "hardwired" into the commons' permissions infrastructure, ensuring that approved uses and users gain access efficiently.[69] Transparency is key: The commons makes clear what is the path to access, what requests have been granted, and what has been learned. Housed in an academic research center, the commons is able to act as an independent "trusted intermediary" for all parties.

The CJDC draws on the "Five Safes" framework to provide data access, balancing the terms and conditions of data providers and the needs of researchers.[70] Under this framework, the CJDC approves only "safe projects" intended to increase knowledge of the civil justice system. Researchers cannot use data for commercial purposes or for administrative purposes that have direct implications for affected individuals in the data. CJDC users are "safe people," who are affiliated with universities, courts, government agencies or non-profit institutions, such as policy institutes and community organizations. The CJDC is a "safe setting" for the data, which can be accessed only on the CJDC's platform. The data are "safe data": Data sharers certify

[66] *See* Grossman, *How Data Commons Can Support Open Science*.

[67] *See* Civ. J. Data Commons, CJDC Governance Model 9 (2021), https://georgetown.app.box.com/s/mj44l32ris28p7i237gci80okkrerf82.

[68] The commons mitigates privacy concerns and promotes uses for the public good by ensuring a user of the CJDC must agree to certain terms, such as accepted and excluded uses of the data, in order to access the data at all. *See* Aziz Z. Huq, *The Public Trust in Data*, 110 Geo. L.J. (forthcoming 2022).

[69] *See* Civ. J. Data Commons, CJDC Governance Model.

[70] *See* Tanvi Desai, Felix Ritchie & Richard Welpton, *Five Safes: Designing Data Access for Research* (U. of the W. of Eng., Working Paper No. 1601, 2016), https://uwe-repository.worktribe.com/output/914745/five-safes-designing-data-access-for-research. The framework has also been adopted as a standard by institutions such as the UK Data Service, *What Is the Five Safes Framework?* UK Data Serv., https://ukdataservice.ac.uk/help/secure-lab/what-is-the-five-safes-framework/; the Australian Institute for Health and Welfare, *The Five Safes Framework*, Australian Inst. of Health & Welfare, https://www.aihw.gov.au/about-our-data/data-governance/the-five-safes-framework; and the Institute for Higher Education Policy, *Five Safes*, Inst. Higher Educ. Pol'y, https://www.ihep.org/wp-content/uploads/2021/02/ihep_fivesafes_onepager_web.pdf.

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

that the data they provide are legally and ethically sourced, and records in the commons are documented so the chain of their provenance is clear. Last, the commons follows the principle of "safe outputs." CJDC users must submit their analytic outputs to the CJDC team, who will ensure that the data were used for the study proposed and that statistical disclosure limitation methods have been applied to prevent re-identification of data subjects.

In essence, the data commons provides a regulated marketplace for data owners and data users. For data owners, the commons facilitates data provisioning and quality control and oversees researcher use of the data. For researchers, the commons provides an interface where researchers can conduct statistical analyses on requested data. The CJDC can enable data discovery across multiple sites and jurisdictions, allowing potential users to browse metadata documentation and propose projects. Project requests are reviewed and approved in accordance with terms of use dictated by the data contributors that govern such questions as who is allowed to use the data, the scope of the research, and the purpose for which the research is undertaken. When projects are approved, CJDC provisions data into a secure workspace with statistical applications that researchers need. By making this information more readily available, knowledge about the prevalence, incidence, and consequences of civil justice problems can be produced, reducing reliance on anecdotal, outdated, or qualitative data alone. CJDC provides a simple, responsive interface that permits researchers to determine what data are available and what research has already been done with various datasets. Users can receive updates about new or refreshed data sources, recommendations on new data topics, and findings from the community.

In its mature form, the CJDC will have a secure infrastructure to automate data provisioning and an interface for data contributors to facilitate data sharing. In addition, it will have separate interfaces for authorized researchers and civil justice institutions so that they have ready access to data for academic and policy research and operational insights. The CJDC will host only machine-readable and documented data, which will be encrypted and held according to industry standards for securely stored data. The CJDC will offer technical assistance or services to help with cleaning, harmonizing, data documentation, and linkages. Funding for the operations and expansion of the Commons might come from federal and state government and foundation grants. The prototype data commons integrates extracts of data from data contributors. A future version of the data commons may have a federated structure as well, enabling the controlled access of data from a data controller's servers. This may alleviate concerns about storage and control.[71]

[71] *See* Huq, *The Public Trust in Data* (proposing a data governance scheme for public trusts under which the data would be subject to the locality's regulation even if data is stored in a server located outside the locality).

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

16.3.2 *Addressing the Barriers*

Institutional Interests and Concerns. As noted above, the greatest barriers to making civil justice data more easily accessible are institutional. Courts are hesitant to share data out of concerns about data security, misuse, and errors, and findings that are embarrassing or ignite unwanted political controversy. The CJDC addresses these concerns by intermediating data sharing and access processes. Court administrators state their terms and conditions for data use, and CJDC screens requests and users, approving only requests that meet these conditions for data access. CJDC also takes over the monitoring role for approved data uses, ensuring scope and data security compliance. By using the well-established Five Safes framework for data access control, the CJDC ensures balance between the concerns of data providers and the needs of researchers. Institutions may also be sensitive to the optics of using these data. The CJDC provides greater privacy protection for data subjects than the open data portals already in place.[72] Finally, the CJDC helps courts avoid surprises by permitting output review. It is critical to have independent research and transparency but equally important for courts to be able to confirm that studies have interpreted local laws or policies soundly. Giving the courts this "heads up" can allow them to frame the findings as opportunities to address issues rather than being surprised by negative headlines.

The CJDC also provides valuable services for data sharers through the production of knowledge to increase their own transparency. We expect that CJDC documentation and data standardization will enable application developers to create tools that work across sites. CJDC data can inform institutions on where things are working well and where they are not, providing evidence to support resource reallocation.

Controlling the Data Disarray. Working with the commons is not an all-or-nothing proposal. Sites can contribute what they have and increase their data sharing over time. For example, a court may share their machine-readable docket information, knowing that key data elements are not yet digitized. Commons users may help the court address this gap, such as through character recognition software or clerical keying. This creates a win-win situation where both the court and researchers gain more usable data. Other courts may share data they know is

[72] One proposed model for how to balance promoting a data public good with protecting privacy is a "safe sharing site." In this governance model, only one party has full access to the raw data and PII. This party then controls what data others can access and analyze, restricting access to the raw data itself. *See* Lisa M. Austin & David Lie, *Safe Sharing Sites*, 94 N.Y.U. L. Rev. 581, 600–601 (2019).

discontinuous, perhaps because of a change in records management vendors, updates to laws or policies, or transitions from paper to electronic records. The CJDC staff will document the data, and CJDC users will enhance detail about the datasets and likely create code that bridges the discontinuity. By fostering a community of users, this knowledge will be shared back to the court and subsequent researchers.

A necessary step in building out the CJDC is the harmonization of court data across jurisdictions, with their disparate data collection policies and case management systems. The CJDC seeks alignment with and adoption of NODS. Important first steps are determining what court data are machine-readable across sites and documenting challenges to NODS standardization within and across datasets. CJDC also standardizes data elements across contributed files, for example, creating a common vocabulary for eviction and debt cases and geocoding location data to facilitate area-level analyses. This standardization can include both manual matching and algorithmic entity resolution, allowing an accurate picture of the *actual* cases despite typos or name changes in the data.

Much of this initial work will continue to be labor- and time-intensive until we understand paths to automate the process of aligning new sites and keeping information up-to-date from courts we have been working with. A key feature of the CJDC is user-friendly metadata (data about the data) noting the sources, time periods available, present or absent fields, and incompleteness rates. As this metadata conforms to NODS, both research and court users will benefit. This metadata goes hand-in-hand with tools to visualize the available data. This data cleaning and sorting is a significant benefit to data contributors, who likely do not have the resources or expertise to perform this work on their own, and provides a provides a clean foundation for research.

Appending Demographic Characteristics. Because court files cannot capture the demographic characteristics of litigants who do not appear in court, linkages to external data sets are necessary to produce this knowledge. The CJDC is pursuing linkages among civil justice, census, and administrative datasets to enable analyses of civil justice eviction and debt-collection cases to develop demographic information about court litigants. These data will be critically important for understanding systemic biases in the aftermath of the pandemic.

16.4 CONCLUSION

Data gaps are not filled overnight, nor by individuals or single institutions. The CJDC is as much a movement as a project. To succeed, it requires participation by courts, practitioners, researchers, policy makers, philanthropists, private industry, and the public. By making civil justice data findable, accessible,

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press

interoperable, reusable, and linked to other data, the CJDC will increase knowledge about biases that distort civil justice processes and outcomes. It can also reveal promising strategies to increase access to civil courts and deepen understanding of how civil justice involvement impacts people living in poverty. If successful, the CJDC can create the basis for evidence-based approaches to civil justice reform.

https://doi.org/10.1017/9781009255301.021 Published online by Cambridge University Press