



ARTICLE

The Necessary and Proper Stewardship of Judicial Data

Zachary D. Clopton & Aziz Z. Huq*

Abstract. Governments and commercial firms create profit and social gain by exploiting large pools of data. One source of valuable data, however, lies in public hands yet remains largely untapped. While the deep reservoirs of data produced by Congress and federal agencies have long been available for public use, the data produced by the federal judiciary is only loosely regulated, imperfectly available to the public at large, and largely ignored by scholars.

The ordinary process of litigation in federal courts generates an enormous volume of data. Especially after recent developments in large language models, this data holds immense potential. It can be used to predict case outcomes or clarify the law in ways that advance legality and judicial access. It can reveal shortfalls in judicial practice and enable the provision of cheaper, better access to justice. It can make legible many otherwise invisible social facts that, if brought to light, can improve public policy. Or the data can fuel private profits, its benefits accruing to a small coterie of data brokering firms capable of monopolizing its commercial use.

This Article is the first to address the complex empirical, legal, and normative questions raised by the untapped public asset of judicial data. It develops a positive, descriptive account of how federal courts produce, dissipate, preserve, or disclose information. This account includes a map of the well-known sources of Article III data (for example, opinions, orders, and briefs), but also extends to a massive volume of “dark data” produced but either lost or buried by the courts. This positive analysis further uncovers a complex administrative framework that erects a plethora of walls and hurdles—some categorical, and some individuated—to slow down or stop public access.

With this positive understanding in hand, we offer a careful analysis of the constitutional questions implicated in decisions to disclose—or to render opaque—judicial data. Drawing attention to the key question of who controls judicial data flows, we demonstrate the existence of sweeping congressional power to regulate judicial data outside of a small zone

* Professors of Law, Northwestern and University of Chicago Law Schools. Thank you to Charlotte Alexander, Jonah Gelbach, William Hubbard, Dan Klerman, Sheri Lewis, Daniel Linna, Peter Martin, Judith Miller, James Pfander, Judith Resnik, Margo Schlanger, David Schwartz, Lesley Zhang, participants at the University of Chicago Works-in-Progress workshop, and our (anonymized) interview subjects.

The Necessary and Proper Stewardship of Judicial Data
76 STAN. L. REV. 893 (2024)

of inherent judicial authority and a handful of instances in which privacy or safety are implicated by disclosure. Congressional authority, therefore, is the rule and not the exception.

Having established these empirical and legal predicates, the Article offers a normative vision of how Congress should regulate the production and dissemination of judicial data in light of the capabilities and incentives of relevant actors. The information produced by the federal courts should not exclusively be a source of private profit for a few data-centered firms. It is a public asset that should be elicited and disseminated in ways that advance the federal courts' mission of equal justice under law.

Table of Contents

Introduction896

I. The Informational Economy of the Federal Courts.....904

 A. The Production of Judicial Data906

 B. The Capture (or Flight) of Judicial Data909

 1. Filings as written records.....909

 2. Transcribing hearings and other interpersonal interactions913

 3. Judicial data loss915

 C. The Disclosure of Judicial Data.....918

 1. PACER.....918

 2. Statistical reports and the integrated database.....921

 3. Commercial databases.....923

 D. The Architecture of Judicial Data: A Summary and Overview925

II. The Constitutional Law of Judicial Data.....927

 A. The Location of Regulatory Authority over Judicial Data.....927

 1. Congressional power over judicial data.....928

 2. Inherent judicial power over judicial data931

 B. Constitutional Arguments for Transparency and Opacity939

 1. Constitutional arguments for transparency939

 2. Constitutional arguments for opacity944

III. Judicial Data as Public Good: An Argument and a Legislative Proposal.....947

 A. Judicial Data as a Public Asset949

 B. Reforming the Stewardship of Judicial Data.....952

 1. Data production and capture952

 2. The public disclosure of judicial data957

 C. The Possibilities of Proper Stewardship963

 1. Non-dispositions and access to justice963

 2. Non-dispositions and judicial performance965

 3. Court data and non-court alternatives.....966

 4. Courts and the social world.....967

Conclusion.....969

Introduction

Governing demands information. Data is caught, yanked, or coaxed from us when we want something from the government or when we fear that it will turn its fearsome coercive powers against us. It can then be filed away in the state's deep archives. When the state acts, it often produces paper trails mapping the entanglements, the bitter conflicts, the gifts bestowed, and the burdens cast upon us.¹ The ensuing troves of government data might be lost, dissipating into air, or kept under wraps, never yielding their secrets. Or they can be zealously stewarded as a public asset that can then be tapped to create public goods for the benefit of all.² So which best describe our present arrangements? And do these arrangements truly align with shared constitutional values?

When it comes to Congress and the executive branch, the rules shaping informational economies are well known. The Constitution's sole disclosure-related provision relates to the transparency of congressional debate.³ By the Third Congress, the systematic publication of congressional records gained a statutory foundation.⁴ Over time, Congress has adopted further laws related to government data, such as the Freedom of Information Act,⁵ the Paperwork Reduction Act,⁶ and the Federal Register Act.⁷ The result is an extensive body of constitutional, statutory, and regulatory provisions channeling legislative⁸

-
1. Recent theoretical work by historians maps the different ways in which state archives have been created and curated. *See, e.g.,* STEWART MOTH, ARCHIVING SOVEREIGNTY: LAW, HISTORY, VIOLENCE 1-22 (2018) (discussing the early historical genealogy of state archives).
 2. *Cf.* Ira S. Rubinstein & Bilyana Petkova, *Governing Privacy in the Datafied City*, 47 *FORDHAM URB. L.J.* 755, 791 (2020) (describing local governments as "data stewards" for their citizens).
 3. U.S. CONST. art. I, § 5, cl. 3.
 4. Act of Mar. 3, 1795, ch. 50, 1 Stat. 443.
 5. Act of July 4, 1966, Pub. L. No. 89-487, 80 Stat. 250 (codified as amended at 5 U.S.C. § 552).
 6. Paperwork Reduction Act of 1995, Pub. L. No. 104-13, 109 Stat. 163 (codified as amended at 35 U.S.C. §§ 3501-3521 and 13 U.S.C. § 91); *see also* 44 U.S.C. § 3502(3)(A) (defining "collection of information" as "the obtaining, causing to be obtained, soliciting, or requiring the disclosure to third parties or the public, of facts or opinions by or for an agency, regardless of form or format").
 7. Federal Register Act, ch. 417, 49 Stat. 500 (1935) (codified as amended at 44 U.S.C. §§ 1501-1511). Congress later amended the Act to create the *Code of Federal Regulations*, which serves as a codification of all current regulations. Act of June 19, 1937, ch. 369, 50 Stat. 304-05 (codified as amended at 44 U.S.C. § 1510).
 8. On early U.S. attitudes toward the relation of publication to democratic norms, *see* Aimée C. Quinn, *Keeping the Citizenry Informed: Early Congressional Printing and 21st Century Information Policy*, 20 *GOV'T INFO. Q.* 281, 281-84 (2003); and Marci A. Hamilton
footnote continued on next page

and executive branch⁹ information—and countless public debates about transparency and opacity in and around both elected branches.¹⁰

When it comes to the Article III judiciary, however, the situation could not be more different. There is a dearth of public debate and scholarship on whether and how federal courts can or should collect, preserve in a lasting form, privately store, or publicly release information about their operation. The legal and administrative infrastructure of judicial data, as a result, largely remains terra incognita for scholars and the general public alike. Instead, public attention tends to focus upon high-profile instances of leaked Supreme Court opinions and private influence campaigns.¹¹ To be sure, these controversies tee up important issues about disclosure, bias, and the epistemic foundations of public confidence in the high court. But they are far from the full picture.

Putting aside the fraught questions raised by the behavior of individual Justices, there is a powerful argument for paying more attention to the ordinary operation of federal judges up and down the judicial hierarchy. Courts nested below the Supreme Court in the federal hierarchy generate an enormous volume of potentially valuable data. In 2021 alone, 526,477 civil and criminal matters were filed in federal district courts—a 24% increase from 2020.¹² In the same year, 326,077 cases were

& Clemens G. Kohnen, *The Jurisprudence of Information Flow: How the Constitution Constructs the Pathways of Information*, 25 CARDOZO L. REV. 267, 280-81 (2003).

9. For example, the Federal Advisory Committee Act requires annual presidential reports on the activities, status, changes, and costs of advisory committees. Federal Advisory Committee Act, Pub. L. No. 92-463, § 6(c), 86 Stat. 770 (1972) (codified as amended at 5 U.S.C. app. § 6(c)); see generally Steven P. Croley & William F. Funk, *The Federal Advisory Committee Act and Good Government*, 14 YALE J. ON REGUL. 451, 458-65 (1997) (discussing its origins). The Freedom of Information Act creates a mechanism for individuals to make information-targeted requests from agencies of the federal government. 5 U.S.C. § 552. The Government Publishing Office must also offer universal online access to statutes and regulations, as well as other government documents and materials. See 44 U.S.C. § 4101(a).
10. Transparency in government processes is the subject of an extensive scholarly literature. See, e.g., Frederick Schauer, Remarks, *The Mixed Blessings of Financial Transparency*, 31 YALE J. ON REGUL. 809, 809-11 (2014); Cary Coglianese, Heather Kilmartin & Evan Mendelson, Report, *Transparency and Public Participation in the Federal Rulemaking Process: Recommendations for the New Administration*, 77 GEO. WASH. L. REV. 924, 924-25 (2009). For a discussion of public debates surrounding government leaks, see David E. Pozen, *The Leaky Leviathan: Why the Government Condemns and Condones Unlawful Disclosures of Information*, 127 HARV. L. REV. 512, 513-21 (2013).
11. See, e.g., Carl Hulse, *Supreme Court Faces Increasing Scrutiny from Congress After Leaks*, N.Y. TIMES (Dec. 7, 2022), <https://perma.cc/7JB6-2EVF>; Adam Liptak, *Alito Says Leak of Ruling Overturning Roe Put Justices' Lives at Risk*, N.Y. TIMES (Oct. 25, 2022), <https://perma.cc/6ZVV-8N8R>.
12. *Federal Judicial Caseload Statistics 2021*, U.S. CTS., <https://perma.cc/9SQH-C9SS> (archived Apr. 29, 2024). These figures were likely shaped by the COVID-19 pandemic. See ROY
footnote continued on next page

resolved.¹³ Meanwhile, there were 46,165 appeals filed in the twelve regional courts of appeals.¹⁴ Every single one of these matters could, in theory, leave behind an informational footprint (although they do not always do so). After all, each case creates a cascade of records, brimming with allegations, facts, deliberations, and decisions. An archive compiling this torrent of “judicial data” could reveal many hidden figures not just about the dynamics of federal adjudication but about American life more generally.

The widely neglected complexity of judicial data extends even further than these simple descriptions would suggest. To be sure, judicial data includes the typical objects of legal research: judicial opinions and perhaps even briefs. But the data of the federal courts involves more than these well-studied documents. Every filing in a legal case—some of which are not publicly available—includes valuable information. Every court hearing—even those that are never transcribed—is a potential font of knowledge. There is a seemingly accepted background norm of transparency: hence the quick and intense outrage at the prospect of Judge Matthew Kacsmaryk holding a nonpublic hearing to adjudicate the legality of the pharmaceutical mifepristone, particularly given the high political stakes of that litigation.¹⁵ This also explains the sharp complaints from commentators of different ideological stripes at the “unprecedented” levels of secrecy during the Google antitrust trial before Judge Amit Mehta in the District Court for the District of Columbia in September of 2023.¹⁶ Even as such deviations from the norm of public hearing are remarked, the routine data loss built into the federal courts’ information system goes unremarked. Yet this uncollected or inaccessible data would permit a multitude of insights were it collected and made accessible. This is the “dark data” of the federal courts.¹⁷

GERMANO, TIMOTHY LAU & KRISTIN GARRI, COVID-19 AND THE U.S. DISTRICT COURTS: AN EMPIRICAL INVESTIGATION 1-2 (2022), <https://perma.cc/6TMR-VHXQ>.

13. *Federal Judicial Caseload Statistics 2021*, *supra* note 12.

14. *Id.*

15. *See, e.g.*, Katie Benner & Pam Belluck, *Judge in Abortion Pill Case Set Hearing but Sought to Delay Telling the Public*, N.Y. TIMES (Mar. 12, 2023), <https://perma.cc/6R72-3FPZ>; Julia Conley, “Unacceptable”: Right-Wing Judge Attempts to Keep Key Abortion Pill Hearing Secret, SALON (Mar. 13, 2023, 1:30 PM EDT), <https://perma.cc/5GWX-VLN3>. *See generally* David S. Cohen, Greer Donley & Rachel Rebouché, *Abortion Pills*, 76 STAN. L. REV. 317 (2024) (discussing the legal landscape surrounding mifepristone).

16. Nico Grant, Cecilia Kang & Tripp Mickle, ‘Unprecedented’ Secrecy in Google Trial as Tech Giants Push to Limit Disclosures, N.Y. TIMES (updated Oct. 4, 2023), <https://perma.cc/WJ34-EXKT>; Thomas Barrabi, *Media Outlets Demand More Access to Secretive Google Antitrust Trial*, N.Y. POST (Oct. 16, 2023, 3:25 PM ET), <https://perma.cc/ZL2A-26P8>.

17. *See generally* DAVID J. HAND, DARK DATA: WHY WHAT YOU DON’T KNOW MATTERS 3-12 (2020) (describing the concept of “dark” or hidden data and identifying the ways in which ignoring its existence can distort empirical analyses).

All of this judicial data could be leveraged for private profit or for the public good. Examination of judicial archives could cast light on many questions of legitimate public concern. For example, when and how the provision of legal counsel is effective, when it is unnecessary, and when it is sorely needed all might be clarified using this data.¹⁸ The metadata created by initial filings, coupled with spatial and temporal data, may generate a better understanding of when and where litigation arises, potentially helping identify barriers to court access.¹⁹ Judicial metadata also has the potential to illuminate the in-court behavior of many important actors, including judges, prosecutors, public defenders and even officials operating behind the scenes.²⁰ This data could help resolve the question, as a team of researchers recently asked, of whether judges are consistent in their evaluation of claims to poverty when they determine in forma pauperis status.²¹ Or we could extend those researchers' agenda to determine the need for standardization of one of the many discretionary decisions federal judges have license to make. Scholars may mobilize this data to explore whether federal judges' sentencing decisions reflect defendants' observed race, ethnicity, or gender.

These research inquiries are just the tip of the iceberg. Court data might also cast indirect light on other questions of public concern beyond the courts. Much of the data generated through federal court litigation, of course, is an imperfect reflection of behavior outside the justice system. Yet examining patterns in court filings over time might illuminate which state and local governments are infringing on civil rights through their policing or municipal court systems. Even when constitutional litigation is absent, the repeated presence of local and state governments as tortfeasors might reveal deeper, more systemic problems in the way that they operate public services or in the lives of the people who use them.²² Corporate litigation may also offer insight into the health or weakness of different sectors of the economy—a warning system that offers insights into sectors heading for a tumble.

18. The question of when legal assistance avails clients was squarely placed on the academic research agenda by 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, 2121-32 (2012) (describing this research agenda and the article's seminal contribution to said agenda).

19. See *infra* Part III.C.1.

20. See, e.g., Kristen Bell, Jenny Hong, Nick McKeown & Catalin Voss, *The Recon Approach: A New Direction for Machine Learning in Criminal Law*, 36 BERKELEY TECH. L.J. 821, 822-26 (2021) (describing new machine-learning tools that are "designed to scrutinize how judges, parole board members, and other decision-makers exercise discretion in the context of criminal law").

21. See Adam R. Pah et al., *How to Build a More Open Justice System: Court Records are Unstructured and Costly to Access—Here's How to Fix It*, 369 SCIENCE 134, 135 (2020).

22. See *infra* Part III.C.4.

The fact that judicial data is so underused is all the more surprising given recent developments in computer science. Large language models (LLMs), such as OpenAI's ChatGPT, Google's Gemini (formerly Bard), and Anthropic's Claude are reinforcement-learning models trained on very large pools of data such as the Common Crawl database.²³ An LLM can be pretrained on a large corpus, and then "fine-tuned" using a smaller, bespoke dataset to offer responses based on the latter.²⁴ Companies such as Hugging Face already make large libraries of pretrained LLMs freely available; these can then be fine-tuned using application-specific data to generate "powerful" results.²⁵ Once trained, an LLM can be used for several different purposes, including text summarization, chatbot interfaces, search, code generation, and essay-like text generation.²⁶

Using judicial data, then, LLMs could allow a user to query a textual corpus generated by federal court litigation to pose almost any imaginable question about the law or other features of judicial practice. What is the legal rule applied in practice to a question of law on which there is no binding precedent? What kind of allegations tend to generate successful certification of class actions? When do judges in a certain circuit tend to admit or reject expert testimony on, say, biological causation or business practices? Do judges differ in how they make such admission decisions between circuits in any specific way? To be sure, early adopters of LLMs among lawyers have quickly identified the risk of errors in generated answers.²⁷ But it seems very likely to us that as protocols for LLM usage become more clearly established and screens for hallucination better designed, the risks of using such instruments to leverage hidden insights and patterns in judicial data will abate. Indeed, the full range of potential uses of LLMs to extract value from judicial data remains unmapped.

Despite the promise of judicial data, there is no comprehensive scholarly treatment of the empirical, legal, and normative questions raised by its

23. For a plain-language summary of the limits and uses of LLMs, see generally Samuel R. Bowman, *Eight Things to Know About Large Language Models*, ARXIV 2304.00612 (2023), <https://perma.cc/294K-C23S>. For summaries of key research papers in the development of LLMs, see Sebastian Raschka, *Understanding Large Language Models: A Cross-Section of the Most Relevant Literature to Get Up to Speed*, AHEAD OF AI (Apr. 16, 2023), <https://perma.cc/VEG3-KPFR>. For a discussion of LLM datasets, see Dennis Layton, *ChatGPT—Show Me the Data Sources*, MEDIUM (Jan. 30, 2023), <https://perma.cc/D3ZL-BYNQ>.

24. *Fine-Tune a Pretrained Model*, HUGGING FACE, <https://perma.cc/JX89-6ZKG> (archived Apr. 17, 2024).

25. *Id.*

26. See Alex Tamkin, Miles Brundage, Jack Clark & Deep Ganguli, *Understanding the Capabilities, Limitations, and Societal Impact of Large Language Models*, ARXIV 2102.02503, at 4 (2021), <https://perma.cc/H4F7-WRJS>.

27. See, e.g., Benjamin Weiser & Nate Schweber, *The ChatGPT Lawyer Explains Himself*, N.Y. TIMES (June 8, 2023), <https://perma.cc/67SK-KD5T>.

existence and potential use. Instead, the smattering of extant articles on judicial data focuses upon the virtues of publishing judicial opinions²⁸ or of opening judicial archives to researchers.²⁹ But these proposals touch upon only a small fraction of the social good that judicial data might foster. Moreover, these suggested reforms do not address the appropriate balance between private gain and public use. And almost nothing has been written on the needful administrative, legislative, and constitutional frameworks surrounding the collection, retention, or distribution of judicial data. As a result, there is simply no account of how judicial data *can* and *should* be lawfully regulated, let alone why it might be regulated and to what ends.

This Article develops a comprehensive descriptive, doctrinal, and normative analysis of judicial data. Because we know so little about how and when judicial data is produced, lost, stored, or distributed, this Article supplements traditional legal research with a small number of targeted interviews with judicial staff and judges,³⁰ primary sources concerning institutional practice and the handling of materials produced in litigation, to draw a positive account of how judicial data arises. It charts how judicial data is produced through a combination of actors internal and external to the court system. Different federal jurisdictions across the United States use the same electronic records systems. However, they vary widely in what they demand of litigants and how they partition employees. As a result, there is surprising variation in front-end data intakes. Courts and judges likewise vary in the care with which information is preserved and the manner in which it is organized.

28. See, e.g., Merritt E. McAlister, *Missing Decisions*, 169 U. PA. L. REV. 1101, 1146-58 (2021); Merritt E. McAlister, *“Downright Indifference”: Examining Unpublished Decisions in the Federal Courts of Appeals*, 118 MICH. L. REV. 533, 561-67 (2020).

29. For example, Lynn LoPucki offers a powerful argument for “the benefits of transparency.” Lynn M. LoPucki, *Court-System Transparency*, 94 IOWA L. REV. 481, 483-513 (2009). In a similar vein, empirically minded scholars have identified classes of opinions that are not available on widely accessible commercial databases operated by Westlaw and Lexis. Elizabeth Y. McCuskey, *Submerged Precedent*, 16 NEV. L.J. 515, 516-19 (2016); Peter W. Martin, *District Court Opinions That Remain Hidden Despite a Long-Standing Congressional Mandate of Transparency—The Result of Judicial Autonomy and Systemic Indifference*, 110 L. LIBR. J. 305, 307-10 (2018). Yet other work considers the effect of new information technologies on the legal system as a whole. An excellent recent example is David Freeman Engstrom & R.J. Vogt, *The New Judicial Governance: Courts, Data, and the Future of Civil Justice*, 72 DEPAUL L. REV. 171, 171-75 (2023). And other scholars have advocated the use of court dockets to supplement opinions when studying judicial decisionmaking. See, e.g., Pauline T. Kim, Margo Schlanger, Christina L. Boyd & Andrew D. Martin, *How Should We Study District Judge Decision-Making?*, 29 WASH. U. J.L. & POL’Y 83, 101-10 (2009).

30. In order to respect the confidentiality of our interview subjects, we identify the interviewees by numbers.

These recording systems, we show, contribute to an information environment in which significant amounts of data can be (and are) lost.³¹

We further identify a small body of administrative practices, largely generated by the Administrative Office of the U.S. Courts, and a small body of statutory laws that both channel the flow of information and decide upon its public dissemination. At present, this legal framework channels public access to judicial data through the Public Access to Court Electronic Records (PACER) system.³² PACER is a government-run database that charges per page for access to court records.³³ Formally, it enables access to almost any specific item of judicial data.³⁴ But its kludge-cluttered and time-consuming interface means that, in practice, public access is tightly constrained and relatively costly. In effect, our study of the administrative and legal structures surrounding federal judicial data reveals the practical importance of commercial firms such as Westlaw (owned by the Thomson Reuters Corporation), Lexis (owned by the RELX Group), and Bloomberg in managing—and hence deciding on—the public flow of information and, in the process, capturing much of its value.

The Article then turns from a positive account of judicial data flows to the doctrinal question of how such data can legally be managed. Who decides, we ask, whether and how judicial data must be closely held or disseminated to the public? The crucial choice here lies between Congress and the courts themselves, exercising what is called an inherent power. Our analysis demonstrates that the text of the Constitution, historical practice in the form of legislation dating back to 1789, and an accumulation of Supreme Court precedent affirm a sweeping congressional authority under Article I to impose binding rules of “practice and procedure” on federal courts, including in relation to judicial data. The general rule of congressional control, however, is subject to important exceptions in specific instances for Article III, Due Process, and Sixth Amendment reasons (among others). Congress, in short, has relatively broad discretion to set the terms on which judicial data is captured and then turned to public use or private gain.³⁵ Congress, however, has strikingly failed to leverage that power through actual legislation.

The descriptive account of how the federal courts handle data and the doctrinal account of who decides how such data is handled lead to the normative question of how such information *should* be managed. Judicial data can generate tremendous social gains by casting light on where courts and other

31. See *infra* Part I.B.3.

32. See *infra* Part I.C.1. (discussing PACER).

33. *Id.*

34. See *id.*

35. See *infra* Part II.A.1.

pivotal institutions are presently falling short.³⁶ Or it can generate large profits for a small handful of “data cartels.”³⁷ We develop a normative case for treating judicial data as a public asset—one that ought to be carefully stewarded in ways that have widespread rather than concentrated benefits. Congress should enact a statute that facilitates the production of such data and then makes it generally available, with narrow exceptions where Article III concerns sounding in judicial autonomy or privacy concerns counsel otherwise. We call upon Congress to take action not only because it has the authority to do so, but also because the courts have, to date, resisted such efforts.³⁸

We do not offer a precise rendition of optimal statutory text. Rather, we develop the normative case for this general approach. Judicial data is generated through the joint action of many individual litigants and public servants.³⁹ It can be leveraged to improve the quality of judicial institutions and to cast light on other important public bodies. Few of these public goods (set to rapidly grow in size due to LLMs) will be realized if judicial data’s dissemination is controlled by a small number of for-profit firms. As a rule, their commercial interests drive them toward limiting the public supply of data through paywalls and user agreements. Treating judicial data as a public asset instead, we argue, resonates with the oath that all federal judges have had to “solemnly swear or affirm” since 1789: that they will “do equal right to the poor and to the rich.”⁴⁰ At the peak of the information age, it would be paradoxical and profoundly unjust to channel the fruits of judicial data toward the wealthy few, at high cost to the many.

Our analysis, in short, has three steps. Part I offers the first systematic account of how judicial data is actually produced and regulated. Part II analyzes the legal and constitutional framework for federal court data. Part III offers a normative argument in favor of a federal statutory reform that would treat judicial data as a public good.

36. See *infra* Part III.C.

37. See generally SARAH LAMDAN, DATA CARTELS: THE COMPANIES THAT CONTROL AND MONOPOLIZE OUR INFORMATION 2 (2022) (introducing the concept of “data cartels”); *id.* at 30 (describing the emergence of “data integration” or “business solutions” firms that control the flow of information generated through digital activity).

38. See *The “Free PACER” Bill Is Not Becoming Law This Year. But That’s Not the Entirety—Nor the End—of the Story*, FIX THE COURT (Dec. 29, 2022), <https://perma.cc/E54W-PM5J> (noting opposition to PACER reforms from the Administrative Office of the U.S. Courts).

39. See *infra* Part I.A.

40. Judiciary Act of 1789, ch. 20, § 8, 1 Stat. 73, 76 (codified as amended at 28 U.S.C. § 453).

I. The Informational Economy of the Federal Courts

A common, often unstated assumption in much legal scholarship is that the federal judiciary is fairly transparent.⁴¹ An exemplar of this view holds that “nearly all documents filed by the parties, orders and other rulings by the presiding judge and the final judgment” are presently available electronically via PACER or a commercial database.⁴² Implicit here is yet another premise: that all relevant actions by litigants and judicial actors are recorded in the first place. A system that makes “all documents” available, but that memorializes only a small fraction of its actions, can hardly be called transparent.

This Part sketches out a far more complex reality. We draw on primary and secondary sources, including interviews with judges, judicial staff, officials in the Federal Judicial Center, and staff of the commercial firms that extract and make available judicial data. The initial interviewees were identified by reaching out to existing contacts in two United States federal courts, as well as scholars who had conducted empirical research using judicial data in the past. In addition, we reached out directly to other judicial personnel and staff at commercial databases. In each interview, we solicited information through an open-ended set of questions focused on a person’s experience working with judicial data. Further, in each interview, we also solicited suggestions about other potential sources and endeavored to expand our sample that way. Our small number of targeted interviews were intended to survey actors with particularized expertise in different aspects of judicial data management. These interviews do not reflect a representative sample in any formal sense, though we endeavored to speak to sources with a range of backgrounds and experiences. We offered anonymity to all of our sources, and where cited, we refer to them by numbers. Full records of all interviews have been shared with the *Stanford Law Review’s* editors.

Working from these sources, we elaborate a bird’s-eye account of how information is collected, stored, and then distributed by federal courts. Our analysis of how judicial data flows in practice yields a series of simple conceptual distinctions for analyzing Article III’s informational economy. These distinctions—between “endogenous” and “external” data, between “hurdles” and “walls,” and between “retail” and “wholesale” disclosure—arise

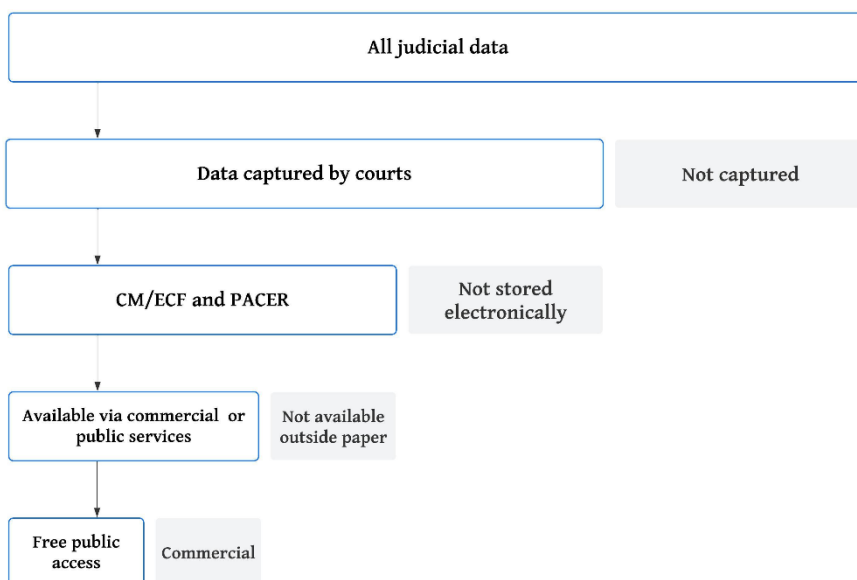
41. We give examples of this view below but emphasize that some scholars have rightly taken more nuanced positions. *See, e.g.,* Kim et al., *supra* note 29, at 96-97 (“[W]hen research aims to understand what factors influence case outcomes or how judges make decisions, an analysis based solely on published opinions is problematic. In particular, such an approach risks producing biased or misleading results because published opinions are not representative of all opinions; opinions are not representative of all district court decisions; and adjudicated cases are not representative of all filed cases.”).

42. Peter W. Martin, *Online Access to Court Records—From Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 855, 855-56 (2008).

from the actual operation of the federal judiciary’s informational economy. They then can be used in the legal and normative analyses of Parts II and III.

At the threshold, we offer a summary of the basic structure of judicial data flow in a graphical form. Figure 1 captures in a very simplified form the stages of data capture and filtration in the federal courts. Each of those stages is isolated and described in more detail in this Part. We also show, in the grey rectangles, the data loss at each stage. The image is not to scale.

Figure 1
Information Flows in the Federal Courts: A Simplified Sketch



Our use of the term “judicial data” also benefits from threshold unpacking. We use this term to describe *all information generated through and about the litigation process in federal courts*.⁴³ It hence reaches “raw” material generated by litigation, such as the nature of a complaint or the number, sequence, and character of filings on a given docket. It also encompasses summary statistics derived from the data, such as simple averages of cases assigned per judge; the median and standard deviation time elapsed from cases’ filing to settlement; and inferred statistics, such as measures of the differences in the voting

43. We include in this definition information acquired through discovery if and when it becomes a court record. *See, e.g.*, FED. R. CIV. P. 56(c) (describing the record for summary judgment).

behavior of circuit judges in ideologically varied panels.⁴⁴ And of course the definition includes data regardless of whether current judicial information systems do or do not collect or publish it. That is, we are sensitive to the possibility of what the statistician David Hand calls “dark data,” or missing data of which no trace or evidence remains.⁴⁵ The definition used here focuses exclusively on federal courts. Of course, it is obvious that state judiciaries also produce similar kinds of data, and must also resolve a parallel cluster of design and distributive decisions. We leave those questions, though, to other work. Hence, references to “courts” and “judges” here should be understood to refer to the federal judiciary.

This Part first addresses the way in which judicial data is produced. We then turn to the way in which it is collected, collated, and shared within the court system. Finally, we discuss the use of disclosure and secrecy, paying particularly close attention to whether information is made available solely to commercial actors or to the public at large.

A. The Production of Judicial Data

The informational systems of the federal courts necessarily begin with the creation of data. This first step has never been carefully studied. In contrast, there is a swell of attention in Congress, the judiciary, and the academy to a thin slice of data-access issues—in particular, around access fees for specific documents.⁴⁶ But all these efforts focus on the final stretch of the information flow. Design decisions about the production and capture of data are essential questions in any regime of information access.

Much data used by the federal courts is produced by parties and their lawyers. Participant-produced information includes original documents created for the litigation process: pleadings, motions, briefs, and the like.⁴⁷

44. See generally Kevin M. Quinn, Essay, *The Academic Study of Decision Making on Multimember Courts*, 100 CALIF. L. REV. 1493, 1497-98 (2012) (describing panel effects in terms of “the extent to which, and the possible reasons why, attributes of a federal appeals court judge’s colleagues on a particular three-judge panel might exert an influence on the judge’s decision in a particular case”).

45. See generally HAND, *supra* note 17, at 3-12 (defining and applying the idea of “dark,” or missing, data).

46. See, e.g., Jonah B. Gelbach, *Free PACER*, in LEGAL TECH AND THE FUTURE OF CIVIL JUSTICE 328 (David Freeman Engstrom ed., 2023) (making the argument that PACER should be free and collecting sources); Nate Raymond, *Making PACER Court Records System Free Wouldn’t Add to Deficit, CBO Says*, REUTERS (Dec. 8, 2022, 4:31 PM PST), <https://perma.cc/JRG9-VG2T> (discussing the proposed Open Courts Act which would make PACER free); *PACER Fees Class Action*, KCC, <https://perma.cc/6WLM-BKC8> (archived Apr. 17, 2024) (collecting settlement documents related to *National Veterans Legal Services Program v. United States*, No. 16-cv-00745 (D.D.C.)).

47. See FED. R. CIV. P. 7 (discussing pleadings, motions, and other papers).

Parties also create information when they complete court forms, which today are typically electronic.⁴⁸ For initial filings, this includes substantial case and party information, including the case type.⁴⁹ For later filings, parties indicate the filing type, any relief requested, and its relationship to any other case documents.⁵⁰ Metadata from these web-based transactions can become relevant, for example, when applying rules about the timely filing of documents.⁵¹ Metadata also could be aggregated and then analyzed to investigate patterns of court access.

Parties (and their lawyers) likewise produce judicial data when non-litigation information—for example, documentary evidence such as business records—is included with another filing. This type of judicial data, as popularized in the movie *Spotlight*, led to a breakthrough in journalists' investigations of child sex abuse by Catholic priests in the Boston area.⁵²

Even the most rudimentary aspects of litigation can produce judicial data of profound importance. Consider what follows when a plaintiff initiates an ordinary civil lawsuit. As all first-year law students know, the plaintiff must file a document called a complaint. The complaint is a litigation document captured as judicial data upon filing. The complaint tells the court what the case is about, why the court has jurisdiction, and what relief is sought.⁵³ If the plaintiff attaches additional documents to the complaint—for example, a medical bill in a personal injury case—those documents become judicial data, too. When parties choose to include such extrinsic documents, they may inadvertently place in the record intimate information that raises serious

48. *Vince v. Rock Cnty.*, 604 F.3d 391, 393 (7th Cir. 2010) (“Virtually every federal court today operates a comprehensive case management system that allows it to maintain electronic case files and offer electronic filing over the Internet.”). Even recently, however, this was not necessarily so. *See, e.g., United States v. Harvey*, 516 F.3d 553, 555–56 (7th Cir. 2008) (“According to the local rules of the Eastern District of Wisconsin, however, the notice of appeal must be filed ‘conventionally on paper.’” (quoting E.D. Wis., Electronic Case Filing Policies and Procedures Manual, Criminal Part III(C) (2006))).

49. *See* Civil Cover Sheet (2024), <https://perma.cc/WQ7Y-JXLP> (also known as a “JS-44” form).

50. *See, e.g., U.S. DIST. CT. S.D. ILL., CM/ECF USER’S MANUAL* § 4.0 (n.d.), <https://perma.cc/T4TF-67M9>.

51. *See, e.g., In re Sands*, 328 B.R. 614, 619 (Bankr. N.D.N.Y. 2005) (using records from the court’s electronic record system to determine whether a bankruptcy-related petition was timely filed).

52. SPOTLIGHT (Participant Media 2015); Jon Henley, *How the Boston Globe Exposed the Abuse Scandal that Rocked the Catholic Church*, *GUARDIAN* (Apr. 21, 2010, 4:00 PM EDT), <https://perma.cc/XPB3-XKLX>.

53. *FED. R. CIV. P.* 8(a).

privacy concerns.⁵⁴ Even at this early stage, therefore, the information flowing through the courts is heterogenous. Some information's disclosure would raise grave privacy concerns; some is less sensitive.

Less well known than the complaint is the "civil cover sheet" that accompanies all complaints.⁵⁵ This is a one-page form that must be filed along with any civil complaint in federal court.⁵⁶ Some jurisdictions require all civil actions to have a civil cover sheet.⁵⁷ Others exempt pro se plaintiffs from this requirement.⁵⁸ In completing the civil cover sheet, a plaintiff (or attorney) answers a series of questions.⁵⁹ The answers are used to categorize the case for recordkeeping purposes. The plaintiff must answer questions about the basis of jurisdiction, the type of relief sought, and the so-called "nature of suit" (NOS) code.⁶⁰ This is a three-digit code that identifies the type of case and is often used to provide aggregate data about the work of the courts.⁶¹ Options for the NOS code are preprinted on the civil cover sheet.⁶²

A comprehensive database of civil cover sheets would be an extremely valuable source of insight into the timing, cyclicity, substance, and distribution of civil litigation in federal courts. A machine-learning instrument or an LLM could analyze this data to generate templates for future litigants—easing one administrative barrier to court access. But no such database exists.⁶³

54. See FED. R. CIV. P. 56(c)(1)(A) (describing evidence to accompany motion for summary judgment).

55. See Civil Cover Sheet, *supra* note 49.

56. *Id.*; see also FED. R. CIV. P. 79(a)(1) (stating that "the Director of the Administrative Office of the United States Courts with the approval of the Judicial Conference of the United States" may prescribe the form and style of the district court's civil docket); ADMIN. OFF. OF THE U.S. CTS., REPORT OF THE PROCEEDINGS OF THE JUDICIAL CONFERENCE OF THE UNITED STATES 50 (1974), <https://perma.cc/ZVD9-CGET>. Congress has created a Judicial Conference of the United States by statute, see 28 U.S.C. § 331, and an Administrative Office of the United States Courts, see 28 U.S.C. §§ 601-613. The latter's director is empowered to "gather, compile, and evaluate all statistical and other information required for the performance of his duties and the duties of the conference with respect to [magistrate judges]." 28 U.S.C. § 604(d)(2).

57. See, e.g., E.D. WIS. CIV. L.R. 3(a).

58. Andrew Hammond, *The Federal Rules of Pro Se Procedure*, 90 FORDHAM L. REV. 2689, 2710 (2022).

59. See Civil Cover Sheet, *supra* note 49.

60. See *id.*

61. See Christina L. Boyd & David A. Hoffman, *The Use and Reliability of Federal Nature of Suit Codes*, 2017 MICH. ST. L. REV. 997, 998-1000 (describing the use of NOS codes and criticizing their reliability).

62. See Civil Cover Sheet, *supra* note 49.

63. We discuss below the federal courts' Integrated Database (IDB), which relies on cover sheets. See *infra* notes 176-84 and accompanying text. But even the IDB's own user
footnote continued on next page

Most of the data discussed in this Subpart so far is *exogenous* in the sense that it is generated by actors who are not directly employed by or acting as agents of the court system. Supplementing—or, better, catalyzed by—these external sources is data generated *endogenously* by actors who are employed by or acting as agents of the court. As with exogenous data, endogenous data can include original documents (such as orders and opinions), designations on the electronic filing system, and metadata produced along the way.

Like other government bodies, a federal court is a “they,” not an “it.” It is not solely composed of judges, but also law clerks, staff of the clerk of the court, staff attorneys, and other employees. Some court-produced data is created by judges. Other, more ministerial data is created by the office of the clerk or other judicial staff. For example, if and when a judge rules on a motion, they may produce an order and opinion. The judge, or else one of the chamber’s staff, will upload these onto the case-filing system with accompanying designations.⁶⁴ Courts have direct control over the quality of endogenous data; they have less control over the quality of exogenous data.

Finally, every in-person court hearing is an opportunity for data production, both with respect to the content and with respect to the outcome of a proceeding. For example, the fact that a hearing occurred can be recorded in an electronic system. The hearing itself may be transcribed. And any outcomes can be memorialized in a written order. Whether any such data is captured in practice is a subject to which we turn below.

B. The Capture (or Flight) of Judicial Data

The fact that judicial data is produced does not mean that it will be retained or made publicly available, let alone leveraged for the public good. In this Subpart, we analyze the dynamics of its preservation or dissipation. We start with the legal and administrative frameworks for creating written judicial data—another surprisingly under-studied domain of law. We then return to the question of how judicial data dissipates.

1. Filings as written records

The statutory history of federal court records began in earnest in the middle of the twentieth century. The 1948 revision of Title 28 of the U.S. Code required that each of the district courts and courts of appeals maintain their

guide acknowledges its shortcomings. See Fed. Jud. Ctr., *The Integrated Database: A Research Guide* 4 (n.d.), <https://perma.cc/KJG6-WREC>.

64. Interview with Source No. 1 (Oct. 18, 2021). All transcripts and notes are on file with the authors.

own records at the court.⁶⁵ A further obligation to preserve court data came two years later in the Federal Records Act (FRA).⁶⁶ A product of the Hoover Commission, the FRA compelled federal agencies to maintain records in cooperation with the National Archives.⁶⁷ Notably, while the courts are often exempt from statutes regulating the administrative state, they do fall within the FRA's scope.⁶⁸ A number of other federal statutes and court rules impose further preservation mandates.⁶⁹

Primary responsibility for implementing these recordkeeping obligations can fall to officials in each court, including court executives or clerks of court.⁷⁰ The Administrative Office of the U.S. Courts—an agency within the federal judiciary that provides centralized support services to the courts⁷¹—offers some support for the decentralized federal judiciary's records management system.⁷² But at the end of the day, “[b]y statute and administrative practice, each court . . . manages [its own] court records.”⁷³

Court filings used to be on paper. Since the 1980s, individual federal courts have allowed some electronic filing of complaints and other materials.⁷⁴ Later, there was a nationwide move to a new system called the Case Management and

65. 28 U.S.C. § 457 (“The records of district courts and of courts of appeals shall be kept at one or more of the places where court is held. Such places shall be designated by the respective courts except when otherwise directed by the judicial council of the circuit.”). The 1948 revision simply restated an existing requirement of 28 U.S.C. § 10 (1940), but added the courts of appeals. *See* Historical and Revision notes to 28 U.S.C. § 457. This requirement was originally imposed by Congress in 1911. *See* Act of Mar. 3, 1911, ch. 231, § 6, 36 Stat. 1087, 1088.

66. Federal Records Act of 1950, ch. 849, § 6, 64 Stat. 578, 583-90 (codified as amended in scattered sections of 44 U.S.C.).

67. RICHARD J. COX, CLOSING AN ERA: HISTORICAL PERSPECTIVES ON MODERN ARCHIVES AND RECORDS MANAGEMENT 3 (2000).

68. The Administrative Procedure Act, for instance, excludes federal courts from its ambit. 5 U.S.C. § 551(1)(B). But the FRA specifies that all parts of the judiciary, except the Supreme Court of the United States, qualify as federal agencies for its purposes. 44 U.S.C. § 2901(14).

69. For court rules, see FED. R. CIV. P. 79; FED. R. CRIM. P. 55; FED. R. BANKR. P. 5003; and FED. R. APP. P. 45(b). For statutes, see, for example, 28 U.S.C. § 753.

70. *See* 10 U.S. CTS., GUIDE TO JUDICIARY POLICY § 610.40 (2019) (on file with authors).

71. *Judicial Administration*, U.S. CTS., <https://perma.cc/C4WZ-GG2X> (archived Apr. 17, 2024).

72. *See* 10 U.S. CTS., *supra* note 70, § 610.30.

73. *Judicial Administration*, *supra* note 71.

74. *See* Robert Anderson, Ellen Borgersen, Paul Brest, William Fenwick & Jefferson Sher, *The Impact of Information Technology on Judicial Administration: A Research Agenda for the Future*, 66 S. CAL. L. REV. 1761, 1766-69 (1993) (documenting technological forms used in the early 1990s); Jason Krause, *100 Innovations in the Law*, ABA J., Apr. 2015, at 34, 36 (“In the late 1980s, electronic terminals for e-filing documents first appeared in federal courts . . .”).

Electronic Case Files, or CM/ECF.⁷⁵ CM/ECF allows parties (or lawyers) and court officials to electronically file the types of documents described above.⁷⁶ Lawyers file pleadings, motions, and briefs. Judges (or staff members) file orders and opinions. Documents are typically filed and stored as PDFs.⁷⁷ When a user files a document in CM/ECF, they must select an event code from a list of options.⁷⁸ These include, for example, “pretrial memorandum” or “amended complaint” or “garnishee’s answers to interrogatories.”⁷⁹ Users can also provide titles or additional text description of discrete items.⁸⁰ These appear on the “docket sheets,” which list all of the documents that have been filed in the case.⁸¹ Certain entries, such as a judge’s minute orders, do not include an underlying document. For these, a text description is all that is captured.⁸²

At first, individual federal district courts had the option of adopting CM/ECF, but were not compelled to do so.⁸³ As late as November 2004, only about half of the district courts had adopted CM/ECF.⁸⁴ Today, however, all courts of appeals, district courts, and bankruptcy courts use CM/ECF.⁸⁵ Judicial data capture, however, is still far from uniform. CM/ECF is not a single system. Instead, it is an overlapping collection of information processing and recording systems maintained by all the individual courts.⁸⁶ As practicing lawyers—to their chagrin—may learn,⁸⁷ each system is set up to require

75. Interview with Source No. 2 (Oct. 29, 2021); *see also* Martin, *supra* note 42, at 861.

76. *FAQs: Case Management/Electronic Case Files (CM/ECF)*, U.S. CTS., <https://perma.cc/7PV6-6XEG> (archived Apr. 17, 2024).

77. *Id.*

78. *See, e.g.*, U.S. DIST. CT. S.D. ILL., *supra* note 50, § 4.0.

79. *See id.* app. D.

80. *See id.* § 4.0.

81. *See* Kendra K. Albert & Mason A. Kortz, Harv. L. Sch. Cyberlaw Clinic, *How to Read a Docket* (2021), <https://perma.cc/B9D4-A3HD>; Roger Michalski, *Transferred Justice: An Empirical Account of Federal Transfers in the Wake of Atlantic Marine*, 53 HOUS. L. REV. 1289, 1296 (2016) (discussing docket sheets).

82. *See* Albert & Kortz, *supra* note 81, at 6 fig.1; *see also* PACER, <https://perma.cc/5J4D-QBPR> (archived Apr. 17, 2024) (allowing users to search for individual cases (“Find a Case”), review the docket sheet, and observe entries with and without associated documents).

83. Interview with Source No. 3 (Oct. 18, 2021).

84. Martin, *supra* note 29, at 320 n.63 (“In November 2004, only fifty-six district courts [out of 94] were using CM/ECF.”).

85. Interview with Source No. 3, *supra* note 83; *see Court CM/ECF Lookup*, PACER, <https://perma.cc/K8F8-VYLU> (archived Apr. 17, 2024).

86. Interview with Source No. 3, *supra* note 83.

87. *Cf.* Debra Cassens Weiss, *Lawyer Who Missed Deadline After Spending 25 Minutes Trying to File ECF Document Gets No Mercy from 10th Circuit*, ABA J. (Dec. 21, 2022, 12:30 PM CST), <https://perma.cc/3V6L-ZPJB>.

separate login credentials.⁸⁸ The federal courts have developed a nationwide CM/ECF login called Next Generation CM/ECF (NextGen) that was aimed to create efficiencies for attorneys and improve digital security, but such a system is far from a unified national document depository.⁸⁹

A recurrent theme in our interviews was that each federal court maintains a surprising degree of freedom with respect to many aspects of CM/ECF design and management.⁹⁰ As a federal judiciary report stated in 2012:

[W]e are currently running some two hundred distinct case management systems in the federal courts, with little consistency as to which release is employed, which features are activated, and what naming conventions are used. The differences are often not apparent and are ill-explained. Our current systems work remarkably well when looked at through the lens of our individual courts; they are maddening to those users who attempt to work in or retrieve information from multiple courts.⁹¹

Here is one example of an interoperability challenge: Although there is a recommended “dictionary” of event codes for all courts, each court maintains its own list of event codes for docket entries.⁹² This means that the same document filed in two courts can receive a different label in the two cases, even when the cases are related. Within each court, individual judges and staffs have substantial freedom as to how cases are docketed. According to one court official, “each chambers [judge and staff] in the district runs its own docket.”⁹³ Endogenously produced data might hence be complete but not amenable to comparison between courts or even between judges.

The CM/ECF system is the most comprehensive archive of judicial data because it is the first repository into which such data is deposited.⁹⁴ But it does not itself provide public access to court records: Neither members of the general public nor litigants themselves can directly access CM/ECF for the

88. See *Attorney Filers for CM/ECF*, PACER, <https://perma.cc/ED6L-R4YN> (archived Apr. 17, 2024) (noting that potential users must “register directly with each district and/or bankruptcy court who uses CurrentGen CM/ECF.”).

89. See *Get Ready for NextGen CM/ECF*, PACER, <https://perma.cc/44UF-BJJS> (archived Apr. 17, 2024).

90. Interview with Source No. 3, *supra* note 83; Interview with Source No. 4 (Nov. 3, 2021); Interview with Source No. 5 (Oct. 28, 2021); Interview with Source No. 1, *supra* note 64.

91. See Martin, *supra* note 29, at 320 (quoting ADMIN. OFF. OF THE U.S. CTS., NEXT GENERATION OF CM/ECF: ADDITIONAL FUNCTIONAL REQUIREMENTS GROUP FINAL REPORT, at iii (2012), <https://perma.cc/W78X-K8JA>).

92. See *supra* note 78 and accompanying text; Interview with Source No. 3, *supra* note 83.

93. Interview with Source No. 4, *supra* note 90. Indeed, at least one interview subject suggested that there was likely substantial variation in the records management *between cases* even before the same judge. Interview with Source No. 1, *supra* note 64.

94. Interview with Source No. 3, *supra* note 83 (noting that PACER pulls its data from the CM/ECF database).

purpose of searching for information or specific documents.⁹⁵ We will turn below to the main access portal, called PACER, which has some search-related capacity.⁹⁶ An institutional separation hence exists between the principal repository of judicial data and the data aggregation actually accessible to the general public.

2. Transcribing hearings and other interpersonal interactions

Not all litigation occurs in writing. District court judges in particular have broad discretion as to whether “to hold hearings or decide an issue on paper.”⁹⁷ Despite the perennially observed death of the trial,⁹⁸ there are still a great many in-person interactions among judges, court personnel, and parties over the course of federal court litigation. These exchanges can generate rich judicial data about why and how officials of the court (including but not limited to judges) interact with litigants and counsel—if, that is, those interactions are ever captured in a transcribed form.⁹⁹

Applicable law governing the means of producing and disseminating hearing transcripts is thin on the ground. Prior to 1948, there was no standardized method to record federal court proceedings; instead, a party wanting a transcript of a hearing would have to hire a private court reporting service.¹⁰⁰ The 1948 Act clarified the mandate to appoint a court reporter in each of the district courts.¹⁰¹ The Act also affirmed district courts must record proceedings in open court, as well as additional proceedings identified by a judge or court rule.¹⁰² Yet the mere fact that the content of a proceeding will be captured by a court reporter does not ensure that the data will be preserved. In civil cases, only “prepared transcripts” must be placed on CM/ECF.¹⁰³ As a result, this is also true for PACER and other public access portals, which depend on CM/EMF for data.

95. *Id.*

96. *See infra* Part I.C.1 (discussing PACER).

97. McCuskey, *supra* note 29, at 536.

98. The locus classicus of this trope is Marc Galanter, *The Vanishing Trial: An Examination of Trials and Related Matters in Federal and State Courts*, 1 J. EMPIRICAL LEGAL STUD. 459, 522 fig.33 (2004) (tracing the declining number of trials in the United States).

99. Interview with Source No. 6 (Nov. 3, 2021).

100. *See* 6 U.S. CTS., GUIDE TO JUDICIARY POLICY § 460.10 (2023), <https://perma.cc/NZF2-DN8G>.

101. Act of June 25, 1948, ch. 646, § 753, 62 Stat. 921 (codified as amended at 28 U.S.C. § 753); *see* 28 U.S.C. § 9a (supp. 5 1940); Historical and Revision notes to 28 U.S.C. § 753.

102. 28 U.S.C. § 753. The resulting transcripts are not protected by copyright. *See* 6 U.S. CTS., *supra* note 100, § 560(a) (citing 17 U.S.C. §§ 101, 105, 506(c)).

103. *See* 6 U.S. CTS., *supra* note 100, § 510.25.10.

Federal statutory law commands that certain parts of the record are transcribed, “including all arraignments, pleas, and proceedings in connection with the imposition of sentence in criminal cases.”¹⁰⁴ But other parts of the record are transcribed only at the request of a judge or party.¹⁰⁵ Judges do not pay for transcript orders, but parties do—unless they are proceeding in forma pauperis or under the Criminal Justice Act.¹⁰⁶

Since 1984, the federal judiciary has, as a matter of policy, also allowed the use of “electronic sound recording” instead of a court reporter.¹⁰⁷ Typically, judges determine whether or not to employ electronic sound recording.¹⁰⁸ Each court is instructed to keep electronic sound recordings, if created, for a prescribed period and to establish a system for public access to the recordings.¹⁰⁹ But recordings are not automatically transcribed or posted to CM/ECF or PACER.¹¹⁰ A transcript is only produced and made available on the docket when requested by a party or judge.¹¹¹ Without a written transcript, no information is recorded about what happened during the proceeding, which lawyers participated, how long the hearing lasted, and how questions of law or fact were resolved.¹¹²

In the courts of appeals, meanwhile, oral arguments are now typically recorded.¹¹³ Individual courts of appeals may give access to recorded arguments. But coverage is far from comprehensive. For example, the Seventh

104. 28 U.S.C. § 753(b); *see also* 6 U.S. CTS., *supra* note 100, § 510.45.10 (citing the statute).

105. 28 U.S.C. § 753(b); *see also* 6 U.S. CTS., *supra* note 100, § 290.20.20(a) (citing the statute).

106. 28 U.S.C. § 753(b), (f); *see also* 6 U.S. CTS., *supra* note 100, § 530.65 (citing the statute). Each district court sets the fee schedule for transcripts, subject to a maximum rate set by the Judicial Conference. 28 U.S.C. § 753(f); *see also* 6 U.S. CTS., *supra* note 100, § 530.10 (citing the statute).

107. 6 U.S. CTS., *supra* note 100, § 350.10; *see also* 28 U.S.C. § 753.

108. 6 U.S. CTS., *supra* note 100, § 310.10(b) (“United States district judges have the discretion to determine the means for recording their proceedings.”). The persistence of human court reporters ought to be surprising given the declining cost of recording and the diffusion of cheap voice recognition tools.

109. *Id.* § 350.40(b), (f).

110. *See Federal Courts Reporting Program*, U.S. CTS., <https://perma.cc/4WLW-3KWM> (last updated Oct. 2, 2023) (describing how and when court reporters will provide written transcripts and noting access limitations to transcripts if they are electronically filed); *compare* 6 U.S. CTS., *supra* note 100, § 350.30-40 (describing preservation of recordings without discussing CM/ECF), *with id.* § 510.25.10 (describing access to filed transcripts via CM/ECF and PACER).

111. *Federal Courts Reporting Program*, *supra* note 110; *see also* 6 U.S. CTS., *supra* note 100, §§ 350.40(g), 510.25.10. According to the Guide to Judiciary Policy, individual courts have worked on integrating electronic sound recording and CM/ECF, but it appears there is no formal judiciary policy on the matter. *Id.* § 3C.50.

112. Interview with Source No. 6, *supra* note 99.

113. Interview with Source No. 2, *supra* note 75.

Circuit provides public access to only select oral argument recordings captured after December 2001.¹¹⁴

3. Judicial data loss

Based on our interviews and study of federal data collection in practice, this Subpart identifies four ways judicial data can be lost.

First, some data is never recorded. If a federal judge decides to hold a colloquy or a hearing without a court reporter or electronic sound equipment, then no data produced during that session will be captured. The statute requiring court reporters applies only to “session[s] of the court” and “other proceeding[s] designated by rule or order of the court,” implying that there are proceedings not so designated—and thus not reported.¹¹⁵ Similarly, a phone call from an attorney to a court clerk; an off-the-record conversation during sidebar; or a judge-supervised settlement conference will not be recorded, even if it determines the proceeding’s outcome.¹¹⁶ Other aspects of litigation will also be lost.¹¹⁷ “Under Federal Rule of Civil Procedure 26(c), the proponent of secrecy must establish good cause for ‘confidential’ or similar designations on a document-by-document basis, and protective orders may not be entered without a particularized finding of good cause.”¹¹⁸ In civil cases, the terms of settlements are typically not recorded unless,¹¹⁹ as in a class action, the judge must authorize the deal.¹²⁰ Work by judges and court staff outside of the

114. See *Public Access to Oral Argument Recordings*, SEVENTH CIR. CT. APP., <https://perma.cc/TM5S-GCA6> (archived Apr. 17, 2024).

115. See 28 U.S.C. § 753(b).

116. In one very unusual case, a district court judge seemed to have instructed his law clerk to run a pretrial conference alone, and then “appear[ed] to have taken steps to ensure that the pretrial conference did not become a matter of record. There [was] no entry on the docket of this case, no Clerk’s Minute Order, that a pretrial conference was ever held.” Parker B. Potter, Jr., *Law Clerks Gone Wild*, 34 SEATTLE U. L. REV. 173, 193 n.101 (2010) (quoting *Sanders v. Union Pac. R.R. Co.*, 193 F.3d 1080, 1083 (9th Cir. 1999) (en banc)).

117. An obvious lacuna is the absence of any record of external influences on judicial actors’ behavior, such as the weather, the time of the hearing, whether the judge had recently eaten, or whether their favored sports team had recently won. Interview with Source No. 7 (Oct. 13, 2021).

118. Karla Gilbride & Jared Placitella, *Overcoming Secrecy*, TRIAL, Nov. 2022, at 18, 19 (footnote omitted).

119. Interview with Source No. 7, *supra* note 117; see also Scott A. Moss, *Illuminating Secrecy: A New Economic Analysis of Confidential Settlements*, 105 MICH. L. REV. 867, 869 (2007) (“Public settlements are the exception, common in only a few types of cases such as class actions and some cases with governmental defendants or plaintiffs.” (footnotes omitted)). Some district courts, however, restrict settlement confidentiality. See, e.g., D.S.C. LOC. CIV. R. 5.03(E) (“No settlement agreement filed with the court shall be sealed pursuant to the terms of this rule.”).

120. FED. R. CIV. P. 23(e).

courtroom, of course, is also unrecorded. And among law clerks, there is a strong norm of confidentiality respecting such communications.¹²¹ Even if clerks have knowledge of what happened in chambers, they are ethically forbidden from reporting that data.¹²²

Second, captured data may not be stored electronically, dramatically reducing the data's utility. Whether and how frequently such loss occurs varies by court: Each court sets its own rules for electronic filing.¹²³ For instance, in the Northern District of Illinois, the charging documents in a criminal case, warrants for arrest, and all documents that require the signature of a criminal defendant are not filed electronically.¹²⁴ In addition, some types of cases are not assigned to CM/ECF in the first place. In the Northern District of Illinois, for example, petty offenses and grand jury matters fall into this category.¹²⁵ To reiterate, if data is not input into CM/ECF, it will not be reproduced in any public-facing system.

Third, there is federal court data that could be recorded on CM/ECF (and hence on PACER), but that will not be captured unless some additional action occurs. As we have explained, transcripts are one such example.¹²⁶ Expert reports, too, are not presumptively filed.¹²⁷ Instead, commercial database providers such as Lexis engage in extensive search efforts to find and store expert reports that are not filed on PACER.¹²⁸ Yet another important category of lost judicial data includes documents filed under seal and redacted portions of documents.¹²⁹

121. See Eugene A. Wright, *Observations of an Appellate Judge: The Use of Law Clerks*, 26 VAND. L. REV. 1179, 1189 n.38 (1973) (“Confidentiality of the work of a judge is an honored tradition among law clerks as well as a strict rule of court.”).

122. See Gregory Bischooping, *Reconceiving Ethics for Judicial Law Clerks*, 12 ST. MARY'S J. LEGAL MALPRACTICE & ETHICS 58, 77 (2021) (“A clerk’s ability to keep confidential judiciary information confidential is just as important post-clerkship as it is during the clerkship.”).

123. Interview with Source No. 1, *supra* note 64.

124. General Order 16-0020: General Order on Electronic Case Filing, at 2 (N.D. Ill. Apr. 20, 2017), <https://perma.cc/6U4T-H78Z>.

125. *Id.*

126. See *supra* notes 103, 107-12 and accompanying text.

127. FED. R. CIV. P. 5(d)(1)(A) (providing that disclosures under Rule 26(a)(1)-(2), among other papers, “must not be filed until they are used in the proceeding or the court orders filing”); FED. R. CIV. P. 26(a)(1)-(2) (listing initial disclosures and expert disclosures).

128. Interview with Source No. 8 (Dec. 13, 2022).

129. Sealing is regulated by precedent and by court rule. See, e.g., 1ST CIR. L.R. 11.0(c)(1), <https://perma.cc/HP79-3CTR>. In general, a district court “may, in its discretion, seal documents if the public’s right of access is outweighed by competing interests.” *Ashcraft v. Conoco, Inc.*, 218 F.3d 288, 302 (4th Cir. 2000) (quoting *In re Knight Publ’g Co.*, 743 F.2d 231, 235 (4th Cir. 1984)). Sealed documents would be on CM/ECF but not PACER. Interview with Source No. 1, *supra* note 64. In addition, some documents are

footnote continued on next page

While unsealing and unredacting are possible, they require additional judicial action.¹³⁰

Fourth, data may be captured in CM/ECF but recorded in ways that render it unusable or unreliable, complicating subsequent retrieval.¹³¹ One example is data rendered unreliable by human error. The CM/ECF system relies on parties, lawyers, and court staff to upload documents, select event codes, and properly identify cases.¹³² Significant rates of error in this process undermine efforts at systematic study. For example, studies have shown high error rates in tagging “opinions.”¹³³ This means that dispositive written opinions may exist on CM/ECF—and yet, as a result of the manner in which they are tagged and labeled, they are in practice exceedingly difficult to locate without *ex ante* knowledge of their existence (or at least *ex ante* reason to look directly at a specific docket).

In addition, many documents loaded into CM/ECF (and so onto PACER) are not text searchable.¹³⁴ This dramatically reduces the usability of such databases.¹³⁵ Because individual courts have discretion in calibrating their CM/ECF systems, data is not coded and memorialized consistently among courts.¹³⁶ For example, “event codes” identifying document types are not standard across different federal courts.¹³⁷

Even for searchable and consistently tagged data, aspects of the CM/ECF system make it suboptimal for capturing or aggregating certain kinds of accurate information. Recall that NOS codes are the primary mechanism for identifying the subject matter of cases¹³⁸—which might be critical to making sense of large volumes of judicial data. The NOS code library is infrequently

also filed as “restricted,” meaning accessible only to judges and parties. Interview with Source No. 2, *supra* note 75.

130. *See* *Nixon v. Warner Commc’ns, Inc.*, 435 U.S. 589, 598 (1978) (recognizing the discretion of a district court judge over unsealing documents).

131. For further thoughts on what accessible judicial data might look like, see Charlotte S. Alexander & Mohammad Javad Feizollahi, *On Dragons, Caves, Teeth, and Claws: Legal Analytics and the Problem of Court Data Access*, in *COMPUTATIONAL LEGAL STUDIES: THE PROMISE AND CHALLENGE OF DATA-DRIVEN LEGAL RESEARCH* 95, 118-21 (Ryan Whalen ed., 2020).

132. *See supra* Parts I.A.-B.

133. McCuskey, *supra* note 29, at 527-30; Martin, *supra* note 29, at 310, 323-25.

134. Interview with Source No. 3, *supra* note 83.

135. *See* Martin, *supra* note 29, at 316-17.

136. *See id.* at 323-25 (discussing individual courts’ discretion to code “written opinions” in compliance with the E-Government Act).

137. Interview with Source No. 2, *supra* note 75; *see, e.g.*, U.S. Bankr. Ct. C.D. Cal., Quick Reference Guide for CM/ECF Users: Frequently Used Docket Event Codes (2022), <https://perma.cc/CNT6-RE8M>.

138. *See supra* notes 60-62 and accompanying text.

updated.¹³⁹ Some are unhelpfully broad; others are narrow (and used infrequently).¹⁴⁰ Many cases might reasonably fit into multiple categories, but filers must select just one classification.¹⁴¹ Big-data research relying on NOS codes alone, then, raises reliability concerns.¹⁴² Similarly, many other case descriptors (for example, whether a petition to proceed in forma pauperis was granted) are not updated during litigation, even though these aspects of cases may change over time.¹⁴³ This further undermines systematic research.

All of this means that the universe of judicial data captured is much smaller (and much less usable) than the universe of judicial data produced.

C. The Disclosure of Judicial Data

Before 1988, judicial data was available only by visiting a courthouse, attending a hearing, or requesting a paper file. Today, those access points have been superseded by electronic forms of storage. It remains possible to attend hearings that are open to the public. But we focus here on the availability of digital judicial data. For most practical purposes, such digital data is of greater value, in part because it can be analyzed using LLMs or similar models.

1. PACER

In 1988, the federal courts launched the program of electronic access that grew into the Public Access to Court Electronic Records (PACER) service.¹⁴⁴ PACER is a national system that extracts data from each court's CM/ECF system, putatively making it available to the public in a single portal.¹⁴⁵ The grant of public electronic access to files was optional until the E-Government

139. Interview with Source No. 3, *supra* note 83; Interview with Source No. 6, *supra* note 99.

140. Interview with Source No. 3, *supra* note 83.

141. Interview with Source No. 6, *supra* note 99. For a more positive view of NOS codes, see, for example, Theodore Eisenberg & Stewart Schwab, *The Reality of Constitutional Tort Litigation*, 72 CORNELL L. REV. 641, 653-54 (1987) ("Relying on the lists probably does not sacrifice much in the way of completeness. . . . Using the Administrative Office lists [based on NOS codes] therefore affords substantial savings in time and money without sacrificing accuracy.").

142. See, e.g., Boyd & Hoffman, *supra* note 61, at 1000 (describing and criticizing the use of NOS codes).

143. See *infra* note 183 (discussing static versus dynamic labelling).

144. See *Nat'l Veterans Legal Servs. Program v. United States*, 968 F.3d 1340, 1344 (Fed. Cir. 2020) ("[T]he Judicial Conference, in 1988, first authorized an 'experimental program of electronic access for the public to court information.'" (quoting ADMIN. OFF. OF THE U.S. CTS., REPORT OF THE PROCEEDINGS OF THE JUDICIAL CONFERENCE OF THE UNITED STATES 83 (1988), <https://perma.cc/3R4P-2XBM>)).

145. Interview with Source No. 1, *supra* note 64; Interview with Source No. 3, *supra* note 83; see generally Gelbach, *supra* note 46 (discussing the PACER database).

Act of 2002.¹⁴⁶ This law required federal courts to provide online access to “docket information for each case”; “the substance of all written opinions”; and “documents filed with the courthouse in electronic form.”¹⁴⁷ According to the Administrative Office, electronic connection of every court to the PACER system in 2009 satisfied the Act’s public-access requirements.¹⁴⁸ PACER extracts information from CM/ECF. But the systems are separate: Lawyers file documents in one place but access them in another.¹⁴⁹

The exact scope of judicial data captured in PACER has evolved over time. Only in September 2007, for example, did the United States Judicial Conference vote to add transcripts to the system.¹⁵⁰ Even for the cases and materials contained on PACER, documents on those dockets that “have not been tagged by the person filing and docketing them” may be difficult or impossible to locate.¹⁵¹

Starting with the Judiciary Appropriations Act of 1991, Congress has authorized the federal courts to collect fees for electronic access to court records.¹⁵² Today, PACER allows users access to court records from bankruptcy, district, and appellate courts at the price of ten cents per page.¹⁵³ To be sure, some federal courts maintain free websites providing access to a subset of documents (for example, opinions and orders of public interest).¹⁵⁴ But PACER is much more comprehensive, is (somewhat) more user-friendly, and cuts across jurisdictions.¹⁵⁵ On PACER, users can search for an individual

146. E-Government Act of 2002, Pub. L. No. 107-347, § 205(a), 116 Stat. 2899, 2913 (codified as amended at 44 U.S.C. § 3501 note).

147. *Id.* § 205(a)(4)-(6).

148. *See Courtwide Compliance with E-Government Act Requirements*, THIRD BRANCH (Admin. Off. of the U.S. Cts., Washington, D.C.), May 2009, <https://perma.cc/67TM-XFEZ>.

149. As noted, though, efforts are underway to provide a single login for both systems via NextGen. *See Get Ready for NextGen CM/ECF*, *supra* note 89.

150. Martin, *supra* note 42, at 855 (citing *Transcripts of Federal Court Proceedings Nationwide to Be Available Online*, U.S. CTS. (Sept. 18, 2007), <https://perma.cc/AFW9-VMDM>).

151. *See* Martin, *supra* note 29, at 323.

152. 10 U.S. CTS., *supra* note 70, § 210.20 (citing Judiciary Appropriations Act of 1991, Pub. L. No. 101-515, § 404, 104 Stat. 2101, 2132-33 (1990) (codified as amended at 28 U.S.C. § 1913 note)). Congress made this authority permanent in the Judiciary Appropriations Act of 1992. *Id.* (citing Judiciary Appropriations Act of 1992, Pub. L. No. 102-140, § 303, 105 Stat. 782, 810 (1991) (codified as amended at 28 U.S.C. § 1913 note)).

153. *See* PACER CASE LOCATOR, <https://perma.cc/79S7-A9QU> (archived Apr. 17, 2024).

154. *See, e.g., Rulings of Special Interest*, U.S. DIST. CT. S.D.N.Y., <https://perma.cc/87MU-ZVHN> (archived Apr. 30, 2024).

155. *See* U.S. Cts., PACER User Manual for CM/ECF Courts 3-4 (2024), <https://perma.cc/5EFL-BMQQ> [hereinafter PACER User Manual]; U.S. Cts., PACER Case Locator (PCL) User Manual 3 (2024), <https://perma.cc/96HX-H32G> [hereinafter PCL User Manual].

case and then access the case's docket sheet.¹⁵⁶ From the docket sheet, users then can select and download individual case documents filed through CM/ECF.¹⁵⁷ Users accessing only a small number of documents are exempt from the per-page fee.¹⁵⁸ Researchers may apply for fee waivers for specific projects.¹⁵⁹ Even with these exceptions, PACER is profitable for the federal courts, costing an estimated \$3 million but raking in some \$150 million per year.¹⁶⁰ That said, the U.S. government recently settled a class action filed in the D.C. District Court over PACER fees, agreeing to refund \$125 million in fees paid by users.¹⁶¹

Even beyond access fees, PACER's architecture implicitly reflects policy choices about what data should be available and to whom. For example, the judiciary has adopted policies for redacting personal identifiers and other sensitive information in case documents, primarily relying on users themselves to flag such information.¹⁶² In contrast to a more centrally regulated system, these policies mean that litigants with more experience of federal court litigation are more likely to exercise control over their disclosures than neophyte litigants. On the other hand, under the Federal Rules of Civil Procedure, the complete records of certain social security and immigration cases are not accessible remotely except to parties and their attorneys. At the same time, members of the public may obtain access in-person at the courthouse.¹⁶³

PACER's search options also channel and limit data access. Its search capabilities are limited. Users may only search a small set of fields.¹⁶⁴ Even PACER's "advanced" case search feature only allows users to limit a search by court type, case name, case number, case type, case status (closed or open), region, NOS code, and dates.¹⁶⁵ Conspicuously absent from this list is the

156. See PCL User Manual, *supra* note 155, at 6; PACER User Manual, *supra* note 155, at 13-15.

157. See PACER User Manual, *supra* note 155, at 15.

158. See PACER CASE LOCATOR, *supra* note 153.

159. See Gelbach, *supra* note 46, at 339-41.

160. *Id.* at 332.

161. See *PACER Fees Class Action*, *supra* note 46 (collecting settlement documents related to *National Veterans Legal Services Program v. United States*, No. 16-cv-00745 (D.D.C.)).

162. See 10 U.S. CTS., *supra* note 70, § 110.20 (collecting privacy policies); see, e.g., *id.* § 330.10.10(b) ("[T]he attorneys in the case are . . . responsible for reviewing [a prepared transcript] for the personal data identifiers required by the federal rules to be redacted, and providing the court reporter or transcriber with a statement of the redactions to be made to comply with the rules.").

163. FED. R. CIV. P. 5.2(c).

164. See PCL User Manual, *supra* note 155, at 6-10.

165. *Id.*

ability to search by judge.¹⁶⁶ Also missing is any ability to conduct full-text searches, let alone natural-language ones.¹⁶⁷ These front-end design choices represent implicit policy decisions on the judiciary's part. One scholar has argued that making the data freely searchable would not face "any real technical feasibility problem,"¹⁶⁸ and yet the courts have elected to keep this data "behind a kludgy and unjustifiably high paywall" with "severely limited search functionality."¹⁶⁹ Of course, marginal cost estimates do not include the costs of creating the PACER system in the first instance. We think that it is likely, however, that PACER fees to date have already covered that outlay.

So, while PACER reflects a radical advance from the pre-digital age, in practice, it is far from optimal. For those without access to a commercial provider of court records, PACER is often the only option. For these typically lower-resource users, just finding the right documents can be a challenge. PACER assumes users are capable of navigating complex search interfaces, thereby privileging tech-savvy litigants and lawyers. PACER then imposes a per-page fee that makes access economically impossible for some. PACER also makes any sort of large-data processing exceedingly cumbersome. PACER's design thus tilts toward narrowly tailored requests by imposing onerous costs on larger requests for higher volumes of documents. This prevents less-resourced users from engaging in wider searches; the latter are available only to the deep-pocketed who are able to absorb the costs of navigating PACER's intricate interface to acquire specific documents. PACER's design, in short, has significant, likely regressive distributional effects and complex policy effects. While case-specific searches are feasible, many larger searches of interest to the public but not of interest to well-resourced users will not occur.

2. Statistical reports and the integrated database

In addition to PACER's archive, the federal courts store and report a limited set of statistical data. Various federal statutes require the judiciary to create statistical reports describing, for example, the annual business of the federal courts¹⁷⁰ or the so-called "Six-Month List" (a list of matters pending before a judge for more than six months).¹⁷¹ To provide these reports, the

166. Martin, *supra* note 29, at 319.

167. *See id.* at 318-19.

168. *See* Gelbach, *supra* note 46, at 348.

169. *Id.* at 329, 347.

170. 28 U.S.C. § 604; *see, e.g., Judicial Business 2023*, U.S. CTS., <https://perma.cc/4N7X-MVEA> (archived Apr. 30, 2024).

171. 28 U.S.C. § 476(a); *see, e.g., September 2023 Civil Justice Reform Act*, U.S. CTS. (updated Sept. 30, 2023), <https://perma.cc/V9PJ-59LZ>. *See generally* Miguel F.P. de Figueiredo, Alexandra D. Lahav & Peter Siegelman, *The Six-Month List and the Unintended*
footnote continued on next page

Administrative Office must rely on designated officials from local courts who “own” the relevant data and are required to provide it by statute or judicial policy.¹⁷² Individual courts, usually through the office of the Clerk of Court, also create additional statistical reports for their own use.¹⁷³ But these reports, and even their existence, are typically not publicized.¹⁷⁴

Statistical reports issued by the Administrative Office and other entities are far from comprehensive. For example, the courts do not provide aggregate information on sentencing that reveals the identity of the sentencing judge or that is searchable by judge.¹⁷⁵

Another way the federal courts package judicial data is through the Federal Judicial Center’s Integrated Database (IDB).¹⁷⁶ Unlike PACER, the IDB allows users to search for categories of cases and download data pulled from reports created by each court.¹⁷⁷ Each case is numerically coded along a number of fields, such as court, NOS, disposition, origin, jurisdiction, pro se status, and class action status.¹⁷⁸ A search query in the IDB produces a list of cases with additional coding, although judge information is not reported.¹⁷⁹

But the IDB is not entirely reliable.¹⁸⁰ For one thing, it depends on case coding known to be error prone.¹⁸¹ So a search for all cases of a particular type does not actually return all cases of that type. Even among returned cases, inter-court variation makes systemic study difficult. One interview subject noted that, at one time, some courts recorded the monetary demand in thousands of dollars while others recorded the full value.¹⁸² There also is a

Consequences of Judicial Accountability, 105 CORNELL L. REV. 363, 365-72 (2020) (explaining the structure and effects of a six-month deadline for certain motions).

172. Interview with Source No. 3, *supra* note 83.

173. *Id.*

174. *Id.*

175. See ADMIN. OFF. OF THE U.S. CTS., REPORT OF THE PROCEEDINGS OF THE JUDICIAL CONFERENCE OF THE UNITED STATES 21-22 (1995), <https://perma.cc/S3D3-345G> (justifying the practice based on “the potential for judge-specific information taken out of context to be misinterpreted”).

176. *Integrated Database (IDB)*, FED. JUD. CTR., <https://perma.cc/6C95-F5CT> (archived Apr. 17, 2024).

177. See Fed. Jud. Ctr., *supra* note 63, at 1.

178. See *id.* at 2; *IDB Civil 1988-present*, FED. JUD. CTR., <https://perma.cc/B9UZ-64VE> (archived Apr. 30, 2024) (to use the database, select “View the live page”).

179. See Fed. Jud. Ctr., *supra* note 63, at 1.

180. Interview with Source No. 6, *supra* note 99 (explaining some of these concerns).

181. *Id.* For extensive discussion of the unreliability and its effects, see generally Gillian K. Hadfield, *Where Have All the Trials Gone? Settlements, Nontrial Adjudications, and Statistical Artifacts in the Changing Disposition of Federal Civil Cases*, 1 J. EMPIRICAL LEGAL STUD. 705 (2004).

182. Interview with Source No. 6, *supra* note 99.

good deal of missing data. For example, imagine a study of federal-question cases in which an important variable was pro se status. Such a study would be challenging using the IDB because more than one million federal-question cases—or one quarter of all such cases—list pro se status as “missing.”¹⁸³ Moreover, if the researcher wanted to include judge identifiers, then each IDB entry would need to be individually connected to a docket sheet on PACER, since the IDB does not record judge names.¹⁸⁴ And the IDB does not include case documents, which are held behind PACER’s paywall.

In sum, the statistical data made available by the federal courts is not comprehensive. This limited data reflects implicit (albeit unstated and controversial) policy judgments. These judgments may track the public-at-large’s interests, or they may embody concerns distinctive to Article III as an institution.

3. Commercial databases

Beyond government databases, commercial entities such as Westlaw, Lexis, and Bloomberg collect and make available judicial data. Though Westlaw and Lexis are likely best known as repositories of judicial opinions, these databases (along with Bloomberg and others) also collect other court data from PACER. Each of these commercial databases pays the same per-page rate as everyone else for PACER.¹⁸⁵ And each database operates in broadly similar ways.

Lexis, for example, connects directly to PACER through an API (i.e., a digital interface).¹⁸⁶ This connection draws two streams of information from PACER. First, it pulls new dockets and updates existing dockets. Second, Lexis conducts bespoke searches for specific information.¹⁸⁷ Its main clients are law firms, non-legal-services corporations, and other participants in the legal market focused on commercial litigation.¹⁸⁸ As a result, the design of the database prioritizes certain types of cases and filings of interest to its users, such as complex commercial cases.

183. See *IDB Civil 1988-present*, *supra* note 178 (to locate, select “3 - federal question” under “Jurisdiction” and “-8 - missing” under “Pro Se”). Pro se status, among other variables, presents a further complication because the variable is static reflecting whether a litigant was pro se or proceeding in forma pauperis at some particular time in the litigation. It is impossible from the IDB to know if a litigant was always or only sometimes in the listed status, and it is impossible to know if the litigant was pro se or not at an identifiable moment in the case (such as when a motion was filed, or an order issued). See *Fed. Jud. Ctr.*, *supra* note 63, at 3-4.

184. *Fed. Jud. Ctr.*, *supra* note 63, at 1.

185. Interview with Source No. 8, *supra* note 128

186. *Id.*

187. *Id.*

188. *Id.*

Having harvested PDF files from PACER, the company transforms these into a searchable text-based form.¹⁸⁹ It also aims to link together documents from the same case (and even to related cases involving the same defendant or plaintiff or lawyer or issue).¹⁹⁰ Like PACER's system of case-specific page interfaces—but more ambitiously—Lexis's database aims to embody relationships in the database that reflect relationships that exist in the larger legal system.¹⁹¹ Strikingly, commercial databases can thereby offer more complex and detailed patterning of federal litigation than PACER. These inquiries, however, entail not just looking at the tags on documents, but also employing lawyers to examine the documents. Even the costly employment of trained lawyers to that end is not perfect, as the docketing data created by judicial personnel can be difficult to classify.¹⁹² Whether LLMs can do a better job remains to be seen.

There are also gaps. Westlaw and Lexis do not have docket sheets or case documents for all federal cases.¹⁹³ Commercial systems presumably have better coverage of data of pecuniary value. But even there, their reach is incomplete. Settlement data is of particular commercial value, for example, but also extremely scarce on PACER.¹⁹⁴ Thus, commercial databases that draw from PACER are also missing commercially valuable data.

Moreover, the search capabilities of these sites, while better than PACER's and the IDB's, are imperfect. Entering the same search in different databases can return dramatically different results. One study found that in response to a simple search in six different databases, only 7% of cases appeared in all six databases and 40% of cases were unique to one database.¹⁹⁵

Finally, and almost self-evidently, these firms are in the business of translating access into profits. Both RELX and Thomson Reuters are amalgams of dozens of smaller entities that capture and supply information about the law, scientific results, news, and firm- and industry-level behavior.¹⁹⁶ Across several of these fields, they extract and monetize data that is the product of other actors' labor, including from competitors that are "either swallowed or

189. *Id.*

190. *Id.*

191. *Id.*; see *Litigation Analytics Tools*, LEXISNEXIS, <https://perma.cc/ZL5S-UQPS> (archived Apr. 30, 2024).

192. Interview with Source No. 8, *supra* note 128.

193. See Christina L. Boyd, Pauline T. Kim & Margo Schlanger, *Mapping the Iceberg: The Impact of Data Sources on the Study of District Courts*, 17 J. EMPIRICAL LEGAL STUD. 466, 472-81 (2020).

194. See *supra* notes 119-20 and accompanying text.

195. Susan Nevelow Mart, *The Algorithm as a Human Artifact: Implications for Legal [Re]Search*, 109 L. LIBR. J. 387, 390 (2017).

196. See LAMDAN, *supra* note 37, at 2-3.

strangled.”¹⁹⁷ As Sarah Lamdan argues, these firms’ oligopolistic control over legal information allows them to “shape the law” by crafting editorial summaries, selecting key numbers and headnotes, and organizing material in some categories rather than others.¹⁹⁸ Such curatorial decisions, however, presumably are driven not by a general consideration of the public good, but rather demand from the firms’ client base.

The basic architecture of commercial databases, therefore, is shaped by the concerns of high-dollar users rather than all potential users of judicial data. The commercial firms creating these databases then charge for access to the information they have collected or created. Highly interested but impecunious litigants are unlikely to benefit from their efforts. This asymmetrical access to information about federal court litigation likely has repercussions for the prospects of future litigation success. That is, well-heeled firms’ disproportionate access to judicial data translates into greater distributional inequalities in the future.

D. The Architecture of Judicial Data: A Summary and Overview

This survey of the informational landscape of the federal courts uncovers several important conceptual dichotomies. These inform key design decisions and so shape the availability and effects of judicial data.

First, judicial data has two kinds of sources. Data can be created by those parties using the courts; we call this “exogenous data.” Alternatively, data can be created by court actors themselves; this is “endogenous data.”¹⁹⁹ Exogenous and endogenous data will vary in predictable ways. They implicate different interests, embody distinct kinds of background information, and can be regulated through different legal and constitutional regimes. The two groups of data producers also have sharply distinct incentives and institutional capacities.

Second, access to data can be impeded by either “walls” or “hurdles.” A wall is a formal or functional barrier to access. When court documents are sealed, for example, a wall prevents public access.²⁰⁰ Functional walls matter, too. If a federal court elects not to capture data, such as an in-chambers conversation, it functionally walls off that data from disclosure.²⁰¹ And even when data is theoretically available, it can be surrounded by a thicket of access hurdles,

197. *Id.* at 5.

198. *Id.* at 75.

199. In our view, even when noncourt actors enter data such as an NOS code using a fixed set of defaults, the resulting data is best characterized as exogenous because it embodies the ultimate judgment of a nonjudicial actor.

200. *See supra* note 129 and accompanying text.

201. *See supra* notes 121-22 and accompanying text.

including PACER fees imposed by the government,²⁰² additional fees demanded by a private firm,²⁰³ or a requirement to show up in person to view the record.²⁰⁴ Formal hurdles include fees for access or requirements that data access be approved. Functional hurdles include various challenges to access. Today, the formats in which data is recorded and the searchability of those formats create hurdles to access.²⁰⁵

Many information policy decisions raise or lower hurdles. When the federal courts release statistical summaries of publicly available data, they are essentially lowering a hurdle for that information. When they choose to omit judge-specific summary statistics, which can be accessed and calculated via PACER only at great cost and difficulty, they are erecting a hurdle around that information.²⁰⁶ In the most extreme case, a hurdle can be practically insurmountable and so function as a wall. In this way, the hurdle design favors certain uses over others because some actors are rendered more likely to gain entry than others.

Third, decisions about walls and hurdles themselves can also take two forms: wholesale and retail. A wholesale access limit occludes an entire class of data. The decision to charge ten cents per page with a three dollar-per-document cap on PACER is a wholesale rule in the sense that it is applied across the board to *all* cases. Retail limits, in contrast, are imposed on a case-by-case basis, often relying on judges or other court officials to make individualized decisions on information policy. The decision to seal or unseal documents in a particular case is retail. Subsidies to overcome hurdles also can be wholesale or retail. The federal courts subsidize access to data wholesale when they provide those statistical summaries; they subsidize access retail when they make available fee waivers on a case-by-case basis.

In sum, there are a host of consequential decisions concerning the production, storage, and dissemination of judicial data. Judicial data is created either by many distributed litigants (exogenously) or by officials paid by the federal government (endogenously). But this labor, much of it funded by the federal government, is captured by a small number of private firms and then distributed to maximize profits rather than to advance the public good. As we argue in Part III, this is not a defensible arrangement.

202. See *supra* note 153 and accompanying text.

203. See *supra* Part I.C.3.

204. See *supra* note 163 and accompanying text.

205. See *supra* notes 131-43 and accompanying text.

206. See *supra* notes 166, 175 and accompanying text.

II. The Constitutional Law of Judicial Data

This Part analyzes the fundamental legal frameworks around judicial data management, in particular how the Constitution assigns authority among different institutional actors to make decisions about judicial data. This question was framed by the Legal Process School in terms of “institutional settlement,” the “who decides” question.²⁰⁷ Here, that means asking whether Congress or the federal courts (either as a collective or acting as individual judges and Justices) has constitutional authority to make such decisions. Our primary focus is on the *constitutional* law of judicial data flows.²⁰⁸ We then consider constitutional arguments for the production and dissemination of judicial data under Article III, the Due Process Clauses, the First Amendment, and the Sixth Amendment. These clauses of the Constitution generate arguments, in effect, for greater transparency, which might bear on how specific pieces or whole categories of information must be withheld by the courts. Finally, we consider arguments running in favor of opacity flowing from Article III and the Due Process Clause.

A. The Location of Regulatory Authority over Judicial Data

A threshold question of law concerns the location of authority to regulate judicial data. Both Article I and Article III contain textual anchors for such regulation. But as we show, it is most plausible to think that Congress has primacy in respect to the regulation of judicial data. However, Congress does not always speak to how judicial data should be handled. Hence, we conclude, judge-made rules operate unless and until ousted by inconsistent legislation. This, indeed, is how in practice Congress and the courts have operated historically—default action by the courts and sporadic, displacing legislative rules.²⁰⁹

The general sweep of Congress’s power over judicial data, however, is also punctuated by some exceptions. A first kind of exception arises when there is an arguable claim of “inherent” power embodied in Article III that cannot be impinged upon by mere statute.²¹⁰ The metes and bounds of such inherent

207. See HENRY M. HART, JR. & ALBERT M. SACKS, *THE LEGAL PROCESS: BASIC PROBLEMS IN THE MAKING AND APPLICATION OF LAW* 4-5 (William N. Eskridge, Jr. & Philip P. Frickey eds., 1994) (“The principle of institutional settlement expresses the judgment that decisions which are the duly arrived at result of duly established procedures [for making decisions] of this kind ought to be accepted as binding upon the whole society unless and until they are duly changed.”).

208. Beyond the jurisdictional statutes that create appropriations legislation funding judicial data flows, Congress has created a Judicial Conference of the United States by statute and an Administrative Office of the United States Courts. See *supra* note 56.

209. See, e.g., 28 U.S.C. § 604 (giving the Director of the Administrative Office broad authority to collect and manage judicial data).

210. See *infra* Part II.A.2.

power have never been clarified. But the most potent constitutional objections to congressional regulation would hone in upon legislative efforts to regulate a judge's decisional process prior to the issuance of a dispositive order or judgment.²¹¹ A second, weaker claim of inherent judicial power might go to matters of procedural and operational judgment necessary to ensure a fair, accurate adjudication.²¹² Nevertheless, the available legal materials suggest that Congress typically has the final say and that most imaginable regulation of judicial data falls outside either zone of autonomous Article III operation.

1. Congressional power over judicial data

As a threshold matter, our focus in this Article is the ordinary labor of district and circuit courts, rather than that of the Supreme Court. The latter has a unique constitutional status: It is the only judicial body expressly required by the Constitution's text,²¹³ and it operates in distinctive ways. For example, there is a rich debate as to whether the Court has a "supervisory" power to establish procedural rules for the lower federal courts.²¹⁴ This kind of power—which concerns an intrabranch relation between different courts—falls largely outside of the scope of this Article.

In contrast to the mandatory status of the Supreme Court, Article III "empowers, but does not require, Congress to create lower federal courts."²¹⁵ Confirming and complementing this power, the horizontal component of the Necessary and Proper Clause of Article I assigns to Congress the power to "make all Laws which shall be necessary and proper for carrying into Execution . . . all other Powers vested by this Constitution in the Government of the United States, or in any Department or Officer thereof."²¹⁶ Since the

211. *See id.*

212. *See id.*

213. U.S. CONST. art. III, § 1 ("The judicial Power of the United States, shall be vested in one supreme Court, and in such inferior Courts as the Congress may from time to time ordain and establish.").

214. Compare Sara Sun Beale, *Reconsidering Supervisory Power in Criminal Cases: Constitutional and Statutory Limits on the Authority of the Federal Courts*, 84 COLUM. L. REV. 1433, 1455 (1984) (offering a broadly positive view that "the supervisory power doctrine has provided the Supreme Court with a flexible tool which it has employed for a number of purposes"), with Amy Coney Barrett, *The Supervisory Power of the Supreme Court*, 106 COLUM. L. REV. 324, 328 (2006) (concluding that "the Constitution's text, structure, and history do not support the proposition that the Supreme Court possesses supervisory power over inferior courts by virtue of its constitutional 'supremacy'").

215. James E. Pfander, Essay, *Federal Supremacy, State Court Inferiority, and the Constitutionality of Jurisdiction-Stripping Legislation*, 101 NW. U. L. REV. 191, 197 & n.25 (2007) (summarizing literature on the Madisonian Compromise).

216. U.S. CONST. art. I, § 8, cl. 18.

foundational ruling in *Sheldon v. Sill*,²¹⁷ the Court has glossed the scope of legislative discretion respecting “inferior federal judiciary” as “plenary.”²¹⁸ Concurrent to the celebrated *Marbury* decision,²¹⁹ the Marshall Court even acquiesced to legislation eliminating so-called “midnight” circuit court judgeships, leaving their former occupants without judicial office.²²⁰

A legislative power to create inferior federal courts *ex nihilo*—and perhaps even to destroy them—does not necessarily imply the lesser power to fashion such tribunals in whatever way Congress deems fit. At a bare minimum, judges of such “inferior courts” (but perhaps not “inferior tribunals”)²²¹ must be guaranteed their offices “during good Behaviour” and must “receive for their Services, a Compensation, which shall not be diminished during their Continuance in Office.”²²² Congress hence could not create an Article III court staffed by judges whose salaries were funded through inconstant annual appropriations or who could be fired without “good cause.” There is, therefore, an irreducible constitutional minimum of institutional content to Article III.

The question here is whether that constitutional floor is defined by the text of Article III’s first section, or whether it is more extensive in relation to informational dynamics. While the text of Article III is inconclusive, there is a long and largely uncontroverted history of legislative direction as to many other margins of lower federal court operation.²²³ There is hence broad agreement that “historical evidence of settlement through nonjudicial practice sometimes figures importantly in constitutional law.”²²⁴ The extensive historical record of legislative control of federal court procedure strongly supports the idea that Congress has power to set authoritative rules for many

217. 49 U.S. (8 How) 441 (1850).

218. John Harrison, *The Power of Congress to Limit the Jurisdiction of Federal Courts and the Text of Article III*, 64 U. CHI. L. REV. 203, 205 (1997) (citing 49 U.S. (8 How) at 448-49).

219. *Marbury v. Madison*, 5 U.S. (1 Cranch) 137 (1803).

220. *See* *Stuart v. Laird*, 5 U.S. (1 Cranch) 299, 308-09 (1803).

221. *See* James E. Pfander, *Article I Tribunals, Article III Courts, and the Judicial Power of the United States*, 118 HARV. L. REV. 643, 672 (2004) (“[Article I] suggests a distinction between the ‘tribunals’ that Congress may ‘constitute’ under Article I and the ‘inferior courts’ that it may ‘ordain and establish’ pursuant to Article III”).

222. U.S. CONST. art. III, § 1. Note that this imposes a constraint on Congress’s exercise of its otherwise plenary appropriations power. That is, creating and filling a judicial office ties Congress’s hands at least in one small way.

223. This history begins with the Judiciary Act of 1789, the Process Act of 1789, and the Crimes Act of 1790. *See* Benjamin H. Barton, *An Article I Theory of the Inherent Powers of the Federal Courts*, 61 CATH. U. L. REV. 1, 34 (2011) (discussing these measures).

224. *See, e.g.*, Richard H. Fallon, Jr., *The Many and Varied Roles of History in Constitutional Adjudication*, 90 NOTRE DAME L. REV. 1753, 1773 (2015).

(albeit not necessarily all) aspects of judicial operation.²²⁵ The Court has repeatedly affirmed that “Article III of the Constitution, augmented by the Necessary and Proper Clause of Article I, § 8, cl. 18, empowers Congress to establish a system of federal district and appellate courts and, impliedly, to establish procedural Rules governing litigation in these courts.”²²⁶

In the 1825 opinion in *Wayman v. Southard*, the Court affirmed Congress’s “power to make laws for carrying into execution all the judgments which the judicial department has power to pronounce.”²²⁷ Almost a century-and-a-half later, in *Hanna v. Plumer*, the Court pointed to “the constitutional provision for a federal court system (augmented by the Necessary and Proper Clause)” as a font of “congressional power to make rules governing the practice and pleading in those courts.”²²⁸ Congress first directed federal courts to incorporate state procedure,²²⁹ and then in the Rules Enabling Act of 1934 authorized the Supreme Court to prescribe uniform rules to govern the “practice and procedure” of the federal district courts and courts of appeals.²³⁰ It is further “plain” to the Court that Congress has power “to prescribe housekeeping rules for federal courts,” such as the rules for transfer between venues.²³¹

225. Section 17 of the Judiciary Act of 1789 allowed federal courts “to make all necessary rules for the orderly conducting [of] business in the . . . courts, provided such rules are not repugnant to the laws of the United States.” Judiciary Act of 1789, ch. 20, § 17, 1 Stat. 73, 83. This was superseded by the Process Act of 1792, which directed federal courts to apply the “modes of proceeding” used in 1789 in the state in which it sat, “subject . . . to such alterations and additions as the courts . . . shall in their discretion, deem expedient, or to such regulations as the Supreme Court . . . shall think proper . . . by rule to prescribe to any circuit or district court.” See Process Act of 1792, ch. 36, § 2, 1 Stat. 275, 276. Congress continued to legislate on procedure throughout the nineteenth century. See, e.g., Conformity Act of 1872, ch. 255, § 5, 17 Stat. 196, 197 (requiring federal district courts to follow state court procedures, except in equity and admiralty matters). In 1934, Congress established the modern form of procedure in the Rules Enabling Act, paving the way for the Federal Rules. Act of June 19, 1934, ch. 651, 48 Stat. 1064 (codified as amended at 28 U.S.C. § 2072).

226. *Burlington N. R.R. Co. v. Woods*, 480 U.S. 1, 5 n.3. (1987).

227. 23 U.S. (10 Wheat.) 1, 22 (1825). The Court rejected an argument that judgment execution was a matter for the states to regulate. *Id.* at 50; cf. *Beers v. Haughton*, 34 U.S. (9 Pet.) 329, 330 (1835).

228. 380 U.S. 460, 472 (1965); see also *Sibbach v. Wilson & Co.*, 312 U.S. 1, 9 (1941) (“Congress has undoubted power to regulate the practice and procedure of federal courts.”).

229. See *supra* note 225.

230. Act of June 19, 1934, ch. 651, 48 Stat. 1064 (codified as amended at 28 U.S.C. § 2072).

231. *Stewart Org., Inc. v. Ricoh Corp.*, 487 U.S. 22, 32 (1988) (quoting *Hanna*, 380 U.S. at 473) (upholding the validity of 28 U.S.C. § 1404(a)).

As described in Part I, some federal statutes do bear directly on judicial data.²³² Nothing in the public domain suggests that these interventions have generated judicial objection. For example, the E-Government Act of 2002 required all federal courts to enable online access “to the substance of all written opinions.”²³³ There is no judicial opinion, or other authority, even hinting that this statutory publication requirement was an improper exercise of congressional authority.²³⁴ To the contrary, in a report for the Advisory Committee on Appellate Rules, then-Judge Samuel Alito praised the E-Government Act as a “solution” for the “unfortunate reality” of a “disparity between litigants who are wealthy and those who are not.”²³⁵ Accordingly, at least in recent history, the congressional authority respecting the *preservation* of dispositive judicial opinions is uncontested and even celebrated by the bench. Some kinds of regulation of judicial data, then, seem well within Congress’s uncontested reach.

In summary, the text of the Constitution, historical practice in the form of legislation dating back to 1789, and an accumulation of Supreme Court precedent affirm a sweeping congressional authority under Article I and Article III to impose binding rules of “practice and procedure” on federal courts.²³⁶

2. Inherent judicial power over judicial data

Judicial power over the creation and disposition of judicial data would necessarily come from the Vesting Clause of Article III, which speaks in definite terms of “the judicial Power.”²³⁷ We begin by distinguishing forms of inherent power, differentiated by whether they can be overridden by Congress.²³⁸ Rather than offering a sweeping, categorical judgment on how to

232. See also *supra* notes 208-09 (noting the powers of the Director of the Administrative Office).

233. E-Government Act of 2002, Pub. L. No. 107-347, § 205(a), 116 Stat. 2899, 2913 (codified as amended at 44 U.S.C. § 3501 note).

234. None of the respondents in our interviews, including judges and judicial personnel, even suggested that the E-Government Act presented any constitutional concern.

235. Memorandum from Samuel A. Alito, Jr., Chair, Advisory Comm. on App. Rules, to David F. Levi, Chair, Standing Comm. on Rules of Prac. & Proc. 6 (May 6, 2005), <https://perma.cc/6AQU-BX7V>.

236. *Sibbach v. Wilson & Co.*, 312 U.S. 1, 9 (1941).

237. U.S. CONST. art. III, § 1. On the importance of the Article III Vesting Clause, see Steven G. Calabresi & Kevin H. Rhodes, *The Structural Constitution: Unitary Executive, Plural Judiciary*, 105 HARV. L. REV. 1153, 1176 (1992) (“The Vesting Clause of Article III must be read as a grant of power; indeed, it appears to be the only *explicit* constitutional source of the federal judiciary’s authority to act.”).

238. Stephen Burbank draws the same distinction “between inherent power in the weak sense (the power to act in the absence of congressional authorization) and inherent power in the strong sense (the power to act in contravention of congressional
footnote continued on next page”

draw this line, we separate two distinct strands of “inherent judicial power” jurisprudence. With respect to each, we consider how and where the line between defeasible and indefeasible judicial power might be drawn (if, indeed, such a line exists) with respect to several issues.

To many, it will likely seem obvious that the federal courts have a measure of authority to set defeasible procedural, evidentiary, and case-management rules. Authority for this (seemingly self-evident) position is hard to find because it is so seldom controverted.²³⁹ In an early treatment of inherent judicial power, Robert Pushaw argued that “judges must have a default power to fill in gaps in the law of procedure and evidence as needed to decide a case.”²⁴⁰ Courts and rule-makers also have suggested that the formal Federal Rules (for example, the Federal Rules of Civil Procedure) and the Rules Enabling Act that authorized them do not oust all inherent powers—implying the existence of some default, judicial rulemaking authority.²⁴¹ Furthermore, the existence of such a default power is consistent with the observed practice of federal courts. Judges have long been in the business of developing and applying interstitial rules and practices related to procedural matters that Congress either has failed to address, or has incompletely canvassed, and where state law cannot be used to fill gaps.²⁴² In many instances, the orderly presentation and resolution of factual and legal questions through litigation would be difficult to imagine without such power.²⁴³ This kind of default

prescription).” Stephen B. Burbank, *Procedure, Politics, and Power: The Role of Congress*, 79 NOTRE DAME L. REV. 1677, 1681 (2004). His weak/strong distinction is parallel to our periphery/core dichotomy elaborated below. See *infra* text accompanying notes 276-79.

239. It is also perhaps hard to square with certain styles of currently dominant originalist analysis.

240. Robert J. Pushaw, Jr., *The Inherent Powers of Federal Courts and the Structural Constitution*, 86 IOWA L. REV. 735, 742 (2001). This is sometimes called a “supervisory” power, although the terminological difference is often inconsequential. See, e.g., *United States v. Blech*, 208 F.R.D. 65, 67-68 (S.D.N.Y. 2002) (declining to exercise the court’s “supervisory power” to quash allegedly abusive witness interviews).

241. See, e.g., FED. R. CIV. P. 16 advisory committee’s note to 1993 amendment (discussing inherent powers in pretrial case management and citing cases).

242. Pushaw, *supra* note 240, at 749-50 (describing federal courts’ creation of procedural and evidentiary rules in areas where Congress was silent and state law could not act as a gap-filler); see also Barton, *supra* note 223, at 32 (“From before the time of the Constitution’s framing until today, courts have had interstitial authority to fill gaps left in congressional acts.”). For enumerations of cases invoking such gap-filling, see Beale, note 214 above, at 1455-64; and Amy Coney Barrett, *Procedural Common Law*, 94 VA. L. REV. 813, 819-32 (2008).

243. See, e.g., *Barker v. Wingo*, 407 U.S. 514, 530 n.29 (1972) (observing that a court may rely on its supervisory powers to “establish[] a fixed time period within which cases must normally be brought”); *Landis v. N. Am. Co.*, 299 U.S. 248, 254 (1936) (asserting that a district court has inherent power to stay proceedings for the sake of “economy of time and effort for itself, for counsel, and for litigants”); *Bowen v. Chase*, 94 U.S. 812, 824

footnote continued on next page

authority is also consistent with the observation that courts' ability to *sua sponte* elaborate procedure rules "does not extend to disregarding a validly enacted and applicable statute or permitting departure from it."²⁴⁴ This authority is hence plausibly ranked an "inherent" judicial power, albeit one subject to displacement by Congress.²⁴⁵

If all such inherent judicial power existed at Congress's whim, the question of institutional settlement would be straightforward. But matters are not so simple. There is also some evidence in the jurisprudence to suggest that the federal courts have an "inherent" power beyond the salary and tenure entitlements enumerated in Article III's text that Congress cannot abrogate.²⁴⁶ But scholars divide on both whether and how much "inherent" judicial power there is beyond the textual incidents of Article III. Glossing Article III's Vesting Clause, for example, William Baude has recently argued that its "vesting of the judicial power is not about the *process* of adjudication."²⁴⁷ Instead, Baude argues, "[I]t refers to the substance of judicial power (which is the power to bind parties and to authorize the deprivation of private rights)."²⁴⁸ If correct, Baude's argument implies that with respect to matters of "process" (including whether judicial data is produced and how it is subsequently handled), Congress has a largely free hand in its choice of regulatory regime.

In contrast, there is both judicial authority and scholarly work that supports that idea of a more extensive conception of inherent "judicial power" than Baude allows. Pushaw, for example, ranges extensively across English legal history to propose that Article III "presupposes a core of authority that stems from the very nature of the judicial office in Anglo-American history."²⁴⁹ As recapitulated by Justice Scalia, this theory homes in on several "elements of that inherent authority [that] are so essential to '[t]he judicial Power,' . . . that they are indefeasible."²⁵⁰ Echoing Pushaw's conclusions, then-Professor Amy Coney Barrett suggested that core inherent judicial power is "not limited to those occasions on which its exercise is absolutely necessary."²⁵¹ In her view, this

(1877) (describing power of courts to consolidate cases "to prevent any practical inconvenience").

244. *United States v. Nat'l City Lines, Inc.*, 334 U.S. 573, 589 (1948).

245. Could Congress eliminate that default rule-setting power? We are skeptical that this is a constitutional or practical possibility.

246. See *infra* text accompanying notes 257-75.

247. William Baude, *Adjudication Outside Article III*, 133 HARV. L. REV. 1511, 1513 (2020).

248. *Id.* at 1513-14. To be clear, Baude is focused on the question of whether adjudication can be performed by non-Article III entities, which is not the question examined here.

249. Pushaw, *supra* note 240, at 844.

250. *Chambers v. NASCO, Inc.*, 501 U.S. 32, 58 (1991) (Scalia, J., dissenting) (citation omitted).

251. Barrett, *supra* note 242, at 882.

power draws on both Article III and the general law and encompasses each court's "inherent authority to regulate its own proceedings."²⁵² Taking an even more extreme posture, David Engdahl argues that Article III protects courts from statutes they deem "detrimental to judicial potency."²⁵³

This broad historical literature has raised a number of other questions, but we focus here on more foundational questions of how far such an inherent power extends and whether it precludes any potential congressional regulation of judicial data.²⁵⁴ That is, for any given exercise of inherent power, where does the zone amenable to congressional regulation end, and where do we enter whatever "small core of inherent procedural authority that Congress cannot reach" exists?²⁵⁵

We think that the easiest way to make inroads into this issue is to disentangle different kinds of inherent judicial power and then ask how to draw the line between core and periphery with respect to each one. Given the spare treatment of inherent judicial power in the Founding moment and early Republic,²⁵⁶ this means looking at post-ratification jurisprudence. Such an approach also has the merit of avoiding strong assumptions about the power of a single, supervening principle of constitutional jurisprudence (selected from several available ones) to resolve all of these issues. For each case, we consider whether (or the extent to which) the "inherent" power in question cannot be ousted by contrary legislative action. We also ask whether the power in question plausibly bites on the regulation of judicial data.

First, the "judicial Power" has been glossed to include the power to "render dispositive judgments."²⁵⁷ The scope of this "judgment power" is uncertain. A cryptic post-Civil War decision called *United States v. Klein* prohibited

252. *Id.* at 882-84.

253. David E. Engdahl, *Intrinsic Limits of Congress' Power Regarding the Judicial Branch*, 1999 BYU L. REV. 75, 163-64.

254. It is, of course, possible that the Supreme Court issues a decision that purports to bind lower federal courts as a matter of its inherent authority. We think that this is sufficiently unlikely, at least where judicial data is concerned, that the question can be bracketed here.

255. Barrett, *supra* note 242, at 816.

256. Pushaw, *supra* note 240, at 822 ("[T]he Convention delegates did not specifically discuss [the] issue [of inherent judicial authority]. Similarly, the Ratification records do not mention inherent power . . ." (footnote omitted)); Barton, *supra* note 223, at 12.

257. *Plaut v. Spendthrift Farm, Inc.*, 514 U.S. 211, 219 (1995) (quoting Frank H. Easterbrook, *Presidential Review*, 40 CASE W. RES. L. REV. 905, 926 (1990)); accord William Baude, *The Judgment Power*, 96 GEO. L.J. 1807, 1815 (2008) ("[T]he judicial power has traditionally been the power to issue binding judgments . . ."). Related to the judgment power is the claim that "[t]he judicial Power' in 'Equity' includes an inherent power to administer a system of equitable remedies that is coextensive with the remedial authority of the English Court of Chancery in 1789." Owen W. Gallogly, *Equity's Constitutional Source*, 132 YALE L.J. 1213, 1221 (2023) (quoting U.S. CONST. art. III, § 2).

legislative “rule[s] of decision[s]” in pending cases—laws whereby “the court is forbidden to give the effect to evidence which, in its own judgment, such evidence should have, and is directed to give it an effect precisely contrary.”²⁵⁸ The Roberts Court, however, has upheld a number of federal statutes singling out specific parties or even ongoing litigation and then de facto directed a judgment for one, known side.²⁵⁹ These cases might be read to imply that even the core adjudicative function can be regulated in surprisingly extensive ways.²⁶⁰ If Congress can enact new statutes during the pendency of a matter with the overt intention of directing a specific outcome in that case, then how much independence is left for Article III actors in respect to more mundane questions of data management?

But even a muscular “judgment power” is easy to square with a broad congressional authority to regulate judicial data. Rules requiring access or requiring secrecy rarely will impinge on the ability to reach authoritative judgments. We can imagine hypotheticals—albeit convoluted ones—in which possible conflict arises.²⁶¹ But it is worth remembering that the reason to consider hypotheticals in the first place is that Congress has never attempted judicial data policies so aggressive that they implicate the judgment power. The absence of legislative practice of this sort may indicate the weakness of Congress’s authority.²⁶² But the force of such arguments from historical absence surely depends on whether the problem Congress has not addressed historically in fact existed in the past. The legislative failure to address a nonexistent problem historically can hardly be probative.

Second, the Supreme Court has repeatedly flagged the “inherent” power of federal courts to manage the litigation process as necessary to ensure a fair process that results in the issuance of binding judgments. The leading cases on this point often use terms suggestive of exclusive judicial authority. In 1821, the Court thus described courts as “vested, by their very creation, with power to impose silence, respect, and decorum, in their presence, and submission to

258. 80 U.S. (13 Wall.) 128, 146-47 (1872).

259. *E.g.*, *Bank Markazi v. Peterson*, 578 U.S. 212, 215, 236 (2016); *Patchak v. Zinke*, 138 S. Ct. 897, 904, 911 (2018).

260. Aziz Z. Huq, *Why Judicial Independence Fails*, 115 NW. U. L. REV. 1055, 1073 (2021) (“The immediate effect of *Bank Markazi* and *Patchak* is to license targeted legislation aimed at the goal of settling specific suits.”).

261. For example, imagine a statute that required judges to record their initial, candid judgments about how a case should ultimately be resolved upon reading a complaint and that mandates that these impressions be recorded in PACER. Judges might plausibly resist on the grounds that the memorialization of such threshold impressions risked distorting their subsequent behavior in the case.

262. In other structural constitutional cases, the Court has suggested that if “earlier Congresses avoided use of [a] highly attractive power, we would have reason to believe that the power was thought not to exist.” *Printz v. United States*, 521 U.S. 898, 905 (1997).

their lawful mandates.”²⁶³ This extends “beyond question” to the “inherent power of a federal court to investigate whether a judgment was obtained by fraud.”²⁶⁴ It has also long been understood to reach the power to punish contempt²⁶⁵ and to sanction counsel,²⁶⁶ even if the court is otherwise without jurisdiction.²⁶⁷ But none of the cases involve a confrontation between a legislative and a judicially created rule. Consequently, it is not clear that the Court was speaking in those cases of a judicial power that Congress could limit, distinct from a core “inherent” power lying beyond its reach.

One strain of this managerial form of inherent judicial power is directly salient to the regulation of judicial data. The Court has justified the existence of inherent powers that are “governed not by rule or statute” by pointing to “the control necessarily vested in courts to manage their own affairs so as to achieve the orderly and expeditious disposition of cases.”²⁶⁸ This thread of inherent power could be read to entail some measure of inherent judicial authority to control the flow of litigation-related documents outside the courthouse.

Perhaps the most relevant authority on this point is the 1978 opinion in *Nixon v. Warner Communications, Inc.*²⁶⁹ The case turned on media access to the twenty-two hours of audio recordings created by President Nixon’s secret microphones in the Oval Office.²⁷⁰ News organizations sought those recordings after the trial of President Nixon’s co-conspirators.²⁷¹ With Justice

263. *Anderson v. Dunn*, 19 U.S. (6 Wheat.) 204, 227 (1821); *accord Ex parte Robinson*, 86 U.S. (19 Wall.) 505, 510 (1874).

264. *Universal Oil Prods. Co. v. Root Refin. Co.*, 328 U.S. 575, 580 (1946). Some of the early cases concerning the inherent equitable power seem to fall under this rubric. *See, e.g., Massie v. Watts*, 10 U.S. (6 Cranch) 148, 160 (1810) (“[I]n a case of fraud, of trust, or of contract, the jurisdiction of a court of chancery [to grant relief] is sustainable . . .”); *see also Gallogly, supra* note 257, at 1272-73 (discussing these early cases).

265. *See, e.g., United States v. Hudson*, 11 U.S. (7 Cranch) 32, 34 (1812) (recognizing such power in dicta); *Young v. United States ex rel. Vuitton et Fils S.A.*, 481 U.S. 787, 795 & n.7 (1987) (collecting cases recognizing inherent judicial authority to punish contempt); *Ex parte Robinson*, 86 U.S. (19 Wall.) at 510 (explaining that since the “power to punish for contempts is inherent in all courts,” the “moment the courts of the United States were called into existence and invested with jurisdiction over any subject, they became possessed of this power”).

266. *See, e.g., In re Snyder*, 472 U.S. 634, 643 (1985) (“Courts have long recognized an inherent authority to suspend or disbar lawyers.”); *Ex parte Wall*, 107 U.S. 265, 273 (1883) (“It is laid down in all the books in which the subject is treated, that a court has power to exercise a summary jurisdiction over its attorneys . . .”).

267. *See, e.g., Cooter & Gell v. Hartmarx Corp.*, 496 U.S. 384, 393-98 (1990).

268. *Link v. Wabash R.R. Co.*, 370 U.S. 626, 630-31 (1962); *accord Chambers v. NASCO, Inc.*, 501 U.S. 32, 43 (1991).

269. 435 U.S. 589 (1978).

270. *Id.* at 591-596.

271. *Id.* at 592, 594.

Powell writing, the Court observed that “[e]very court has supervisory power over its own records and files, and access has been denied where court files might have become a vehicle for improper purposes,” such as “to gratify private spite or promote public scandal,” for instance, through the publication of “the painful and sometimes disgusting details of a divorce case.”²⁷² The Court did not have to determine how this “supervisory” power ought to be exercised. In the Presidential Records Act,²⁷³ Congress had already imposed “an administrative procedure for processing and releasing to the public, on terms meeting with congressional approval, all of petitioner’s Presidential materials of historical interest, including recordings of the conversations at issue.”²⁷⁴ This language implies that the supervisory power at issue in *Warner Communications* may have been subject to congressional override. The decision does not address whether the referenced inherent power is *always* subject to congressional regulation, and the Court has not revisited the issue since. In other instances, lower courts applying *Warner Communications* have used a balancing test to determine whether “judicial records” should be disclosed or sealed from public view.²⁷⁵

In our view, the best way of reconciling these legal materials is as follows: There is a large “periphery” of legal authority that federal courts exercise in crafting procedural rules, managing their dockets and courtrooms, and attending to the larger institutional needs of the court. When courts generate internal practices with respect to the handling of judicial data, these are *pro tanto* valid absent statutory authority for the reasons Pushaw suggests.²⁷⁶ But such procedures can also be ousted by a contrary legislative instruction. This is so, as *Warner Communications* shows, even when a court might otherwise have a plausible claim to “supervisory” power.²⁷⁷ In addition, there is likely some “core” of nondefeasible inherent judicial power under Article III beyond tenure and salary protections, but it is very modest in scope.

272. *Id.* at 598 (quoting *In re Caswell*, 29 A. 259, 259 (R.I. 1893)). For another example of a case in which access to judicial filings was denied because of the “spiteful” motive of the requester, see *Mirlis v. Greer*, 952 F.3d 51, 56 (2d Cir. 2020).

273. Presidential Records Act of 1978, Pub. L. No. 95-591, 92 Stat. 2523 (codified as amended at 44 U.S.C.).

274. *Warner Commc’ns*, 435 U.S. at 603.

275. See, e.g., *Mirlis*, 952 F.3d at 55-56; *Metlife, Inc. v. Fin. Stability Oversight Council*, 865 F.3d 661, 665-66 (D.C. Cir. 2017).

276. See Pushaw, *supra* note 240, at 744-47.

277. Indeed, the *Warner Communications* Court was careful to stress that it was “addressing only the application in this case of the common-law right of access to judicial records” and did “not presume to decide any issues as to the proper exercise of the Administrator’s independent duty under the statutory standards.” 435 U.S. at 607.

The available legal materials suggest that this “core” of nondefeasible inherent power that pertains to judicial data is small indeed. As noted, the “judgment power” seems to have little bite with respect to judicial data. It is hard to see how, even indirectly, data-management rules could facially undermine the power to render dispositive judgments.²⁷⁸ An inherent judicial power to regulate the litigation process, if one exists, might capture some data-management issues. Its scope might be indexed by the judge’s institutional interest in managing a litigation process free of fraud, malicious deployments of litigation-related powers, and the like. If the contempt power is inherent, for example, then Congress might not be able to prohibit federal courts from using it to enforce compliance with duly issued data-production rules. If *Warner Communications* were taken to be a statement of core inherent authority, then it might support a wider inherent power over court records. But even that decision deferred to legislative preferences as embodied in an act of Congress.²⁷⁹ Again, we also think the “core” inherent authority over data is unlikely to be contested often in practice because it is relatively narrow in scope. The available evidence, in short, suggests that there is ample space for Congress to step in to control the flow of judicial data in ways that do not fundamentally threaten courts’ ability to function or impinge on their inherent authority.

Finally, if Congress did attempt to intrude on core inherent authority, it would be glib to assume that *any* conflict would lead to invalidation of a statutory command. More subtle reconciliations are also imaginable. For example, a court confronted by an apparent conflict might toggle among different ways of defining the competing institutional concerns in an effort to bring them into reconciliation, as in *Warner Communications*.²⁸⁰ Indeed, a court may be obligated to attempt such a reconciliation. Think of this as an “avoidance canon” cautioning against the creation of interbranch conflict. Interpreting the Suspension Clause of Article I, for example, the Court has allowed non-Article III courts to function as substitutes for ordinary federal court review if they meet a threshold of procedural regularity and thoroughness.²⁸¹ By analogy, a federal statute regulating even a core

278. We can imagine as-applied challenges, but these would turn on far-fetched factual hypotheses.

279. See *Warner Commc’ns*, 435 U.S. at 603-07.

280. *Id.* at 607-08 (underscoring the narrowness of its ruling in light of the complex institutional stakes of the case).

281. See, e.g., *Swain v. Pressley*, 430 U.S. 372, 381-84 (1977) (approving of an alternative to habeas corpus in a non-Article III court). In another case, the Court declined to find a statutory procedure for shifting habeas proceedings between different Article III forums “inadequate or ineffective” per se, and therefore declined to reach constitutional questions. *United States v. Hayman*, 342 U.S. 205, 219-23 (1952). Both of these were relatively easy cases in the sense that “[t]he statutes at issue were attempts to streamline

footnote continued on next page

production or handling of judicial data might be valid, provided it adequately addressed Article III concerns.²⁸²

B. Constitutional Arguments for Transparency and Opacity

Congress, in short, has wide discretion to decide how judicial data is handled. But this power is subject to exceptions. In the following two Subparts, we consider arguments that the Constitution constrains the options for managing judicial data. Our aim here is not to advocate for particular legislative reforms. It is rather to map the space of potential constitutional arguments pushing or pulling in different directions.

An illuminating way to organize this discussion is around a simple distinction between transparency claims and their inverse, opacity claims. On the transparency side, there are First and Sixth Amendment, Due Process, and Article III interests. In the opacity column are property and privacy interests—which receive constitutional protection under the Due Process Clauses—and again, Article III interests. Of course, not all of these constitutional interests are implicated in each and every dispute over judicial information. Nevertheless, we work through this range of arguments to show the potential constitutional stakes of judicial data.

1. Constitutional arguments for transparency

To begin with, the Supreme Court has recognized a right of public access to some judicial data. Depending on the circumstances in which it is invoked, this right can be grounded in a common-law right of access to government records, or (as a supplement or an alternative) the First and Sixth Amendments to the Constitution.²⁸³ It is analytically useful to separate out these common-law and constitutional strands. In practice, however, common-law and

habeas corpus relief, not to cut it back.” See *Boumediene v. Bush*, 553 U.S. 723, 774 (2008). *Boumediene* turned on a more complex question in which a habeas substitute was plainly worse in some respects (due process) and better in others (preventing improper disclosures). See *id.* at 776, 787-92, 796.

282. *Warner Communications’s* reliance on the Presidential Records Act as an adequate substitute for the common-law right of access is also somewhat analogous. See 435 U.S. at 602-08. The statute there, however, was a substitute for the common-law right of access to judicial documents and not for the Article III power to manage judicial data. See *id.* Still, the case is at least suggestive.

283. For examples of the Court recognizing this right, see *Warner Communications*, 435 U.S. at 597-98 (acknowledging a “common-law right of access to judicial records” but noting that this right is “not absolute”); and *Bank of America National Trust and Savings Association v. Hotel Rittenhouse Associates*, 800 F.2d 339, 344 (3d Cir. 1986) (stating that “the right of access is firmly entrenched . . . whether grounded on the common law or the First Amendment,” but “not absolute”).

constitutional arguments are blended in ways that makes their separate treatment challenging. An example of this blending is found in *Warner Communications*, which framed the scope of inherent Article III power in terms of the common-law “right to inspect and copy public records and documents, including judicial records and documents.”²⁸⁴ This common-law right, which has been traced back to the fourteenth century by earlier scholars,²⁸⁵ has been invoked to challenge the sealing of briefs and attached materials.²⁸⁶ It is, however, limited to “judicial records.”²⁸⁷ Moreover, it is unclear whether this right only prohibits walls or whether it also precludes the establishment of hurdles (including financial ones) as against public access to data.

The common law right runs alongside, and often intertwines with, two constitutional claims. One is the First Amendment “guarantees of speech and press” that “prohibit government from summarily closing courtroom doors which had long been open to the public at the time that Amendment was adopted,” at least in criminal cases.²⁸⁸ Prior restraints on the release of trial-related material are disfavored.²⁸⁹ Circuit courts have extended these First Amendment principles to other documents produced in criminal proceedings.²⁹⁰ The Sixth Amendment right to public trials in criminal cases

284. *Warner Commc’ns*, 435 U.S. at 597 (footnote omitted).

285. Note, *The Common Law Right to Inspect and Copy Judicial Records: In Camera or On Camera*, 16 GA. L. REV. 659, 660-61 (1982).

286. *E.g.*, *Metlife, Inc. v. Fin. Stability Oversight Council*, 865 F.3d 661, 665-66 (D.C. Cir. 2017).

287. For an application of this vague concept to videotapes of President Clinton’s testimony in a criminal case related to the Whitewater scandal, see *United States v. McDougal*, 103 F.3d 651, 652, 657 (8th Cir. 1996) (“Although the public had a right to hear and observe the testimony at the time and in the manner it was delivered to the jury in the courtroom, we hold that there was, and is, no additional common law right to obtain, for purposes of copying, the electronic recording of that testimony . . . [T]he videotape itself is not a judicial record to which the common law right of public access attaches.”).

288. *Richmond Newspapers, Inc. v. Virginia*, 448 U.S. 555, 576 (1980); *see also* *Globe Newspaper Co. v. Superior Ct.*, 457 U.S. 596, 610-11 (1982) (invalidating a Massachusetts statute barring public access to certain trials); *Press-Enter. Co. v. Superior Ct.*, 464 U.S. 501, 503, 508, 509 n.8, 513 (1984) (extending the right of access access to voir dire); *id.* at 516-20 (Stevens, J., concurring) (elaborating on the First Amendment as the source of the right of access). The public, no less than the press, benefits from these rights. *Cf.* *Saxbe v. Wash. Post Co.*, 417 U.S. 843, 850 (1974) (“[N]ewsmen have no constitutional right of access to prisons or their inmates beyond that afforded the general public.”) (quoting *Pell v. Procunier*, 417 U.S. 817, 834 (1974)).

289. *Neb. Press Ass’n v. Stuart*, 427 U.S. 539, 569-70 (1976).

290. *See, e.g.*, *Seattle Times Co. v. U.S. Dist. Ct.*, 845 F.2d 1513, 1515-17 (9th Cir. 1988) (pretrial release proceedings and documents); *United States v. Haller*, 837 F.2d 84, 86 (2d Cir. 1988) (plea hearings and plea agreements).

also informs the reasoning of some of these opinions.²⁹¹ But circuit courts have understood similar principles to apply broadly in civil litigation as well.²⁹² One district court has extended this logic to cast doubt on whether a state court's bar on scraping judicial databases is consistent with the First Amendment.²⁹³

Second, the Due Process Clauses require a judge who is "impartial in weighing the evidence presented before him."²⁹⁴ The right to an impartial judge might be thought to imply some right to access material that might show bias.²⁹⁵ But as we have seen, current practice is to make it difficult to secure judge-specific information. Nor is there directly relevant case law on the question. In one unusual case involving a state court judge who had already been convicted for taking bribes, the Supreme Court used a "good cause" standard to gauge whether discovery should be allowed in a federal habeas challenge to a state criminal conviction.²⁹⁶ This standard has not been fleshed out, however, for ordinary criminal or civil litigation.

There is reason to be skeptical that a federal court will recognize a constitutional right to scrutinize the bona fides of judicial action. In the case of prosecutors and officials exercising prosecutorial discretion in parallel contexts, for example, there is a high bar for discovery. It is usually difficult to secure documentary evidence of improper motive.²⁹⁷ The Court's recent shift to an "objective"²⁹⁸ standard of judicial impartiality for Due Process purposes in civil cases means that there is less pressure to open a court's files under constitutional compulsion in comparison to a standard that focused on good-

291. See, e.g., *Richmond Newspapers*, 488 U.S. at 573-75; cf. *In re Oliver*, 333 U.S. 257, 266-270 (1948) (describing the Sixth Amendment's public trial right as a "safeguard against any attempt to employ our courts as instruments of persecution").

292. E.g., *N. Jersey Media Grp., Inc. v. Ashcroft*, 308 F.3d 198, 204-05 (3d Cir. 2002) (recognizing a right of access outside of the criminal context but denying media access to deportation proceedings).

293. *S.C. State Conf. of the NAACP v. Kohn*, No. 22-01007, 2023 WL 144447, at *1, *8 (D.S.C. Jan. 10, 2023).

294. *In re Murchison*, 349 U.S. 133, 138 (1955); see also *Aetna Life Ins. Co. v. Lavoie*, 475 U.S. 813, 820-21, 825 (1986); *Tumey v. Ohio*, 273 U.S. 510, 523 (1927); *Marshall v. Jerrico, Inc.*, 446 U.S. 238, 242-44 (1980) (discussing several of these cases in a Fifth Amendment challenge).

295. See, e.g., *Rippo v. Baker*, 580 U.S. 285, 286-87 (2017) (per curiam) (citing *Aetna*, 475 U.S. at 825) (discussing the legal standard used to determine whether a litigant is "entitled to discovery or an evidentiary hearing" respecting judicial bias).

296. *Bracy v. Gramley*, 520 U.S. 899, 900-03, 908 (1997) (applying the "good cause" language of FED. HABEAS CORPUS R. 6(a)).

297. See Ann Woolhandler, Jonathan Remy Nash & Michael G. Collins, *Bad Faith Prosecution*, 109 VA. L. REV. 835, 848 (2023) ("When motions are brought to dismiss a pending prosecution, the motions will often fail when a defendant seeks discovery from prosecutors."); see generally *id.* at 848-50 (citing cases).

298. *Caperton v. A.T. Massey Coal Co.*, 556 U.S. 868, 879 (2009).

faith perceptions of partiality. It seems likely that judges will show greater concern over the confidentiality of judicial deliberations as opposed to prosecutorial judgment as a simple matter of institutional self-regard. That is, judges are more likely to be more sympathetic to concerns about the integrity of judicial rather than prosecutorial decisionmaking.

Finally, it is possible to imagine an Article III argument for transparency that is not identified or discussed in the case law. The issue would arise if Congress prohibited publication of information that a court would prefer to disclose. This argument would focus on the interest the judge has in securing the litigants' and the public's acceptance of judgments.²⁹⁹ Alternatively, it might be premised on the idea that a decision supported by publicly available reasons is more legitimate because it will be more likely to track empirical truths and legal norms.³⁰⁰ Reasoning in this vein, judges,³⁰¹ political theorists,³⁰² and legal scholars³⁰³ have underscored the legitimizing role played by a judge's proffer of justifications. This interest in transparency that is advanced by the publication of an opinion is compatible with an Article III interest in opacity to deliberations: Both can be understood to serve the large goal of legitimating judicial decisionmaking. The transparency interest, however, is implicated only when a judge is precluded from effectively communicating the basis for their decision to the litigants and to the public.³⁰⁴

299. *Cf. id.* at 889 (describing public confidence in the courts as “a vital state interest” but not defining what counts as “public”).

300. *Cf. Frederick Schauer, Giving Reasons*, 47 STAN. L. REV. 633, 657 (1995) (“[W]hen institutional designers have grounds for believing that decisions will systematically be the product of bias, self-interest, insufficient reflection, or simply excess haste, requiring decisionmakers to give reasons may counteract some of these tendencies . . . [T]he very time required to give reasons may reduce excess haste and thus produce better decisions.”).

301. *See* Ruth Bader Ginsburg, Lecture, *The Obligation to Reason Why*, 37 U. FLA. L. REV. 205, 221 (1985) (“A judgment expressing no reasons presents the appearance of arbitrariness.”).

302. *See* JOHN RAWLS, POLITICAL LIBERALISM § 6 (expanded ed. 2005) (describing “public reason” as the essential function of courts in a constitutional regime); *see also* John Rawls, *The Idea of Public Reason Revisited*, 64 U. CHI. L. REV. 765, 767 (1997) (describing “the discourse of judges in their decisions, and especially of the judges of a supreme court” as a core part of the “public political forum”).

303. *See, e.g.,* Mathilde Cohen, *Sincerity and Reason-Giving: When May Legal Decision Makers Lie?*, 59 DEPAUL L. REV. 1091, 1143 (2010) (“Reason-giving is an efficient tool for supervision within the judicial hierarchy.”).

304. The Foreign Intelligence Surveillance Court (FISC), however, has rejected the idea that there is a First Amendment interest in the publication of its otherwise classified and undisclosed opinions. *In re Ops. & Ords. of this Ct. Addressing Bulk Collection of Data Under the Foreign Intel. Surveillance Act*, No. 13-08, 2020 WL 897659, at *1, *9 (FISA Ct. Feb. 11, 2020). This has practical effects on the ability of potentially affected parties to appeal the court's orders. *See* Nola K. Breglio, Note, *Leaving FISA Behind: The Need to Return to Warrantless Foreign Intelligence Surveillance*, 113 YALE L.J. 179, 189 (2003)

footnote continued on next page

The ideal of transparency has been critiqued of late for its “neoliberal” tendencies.³⁰⁵ But this critique lacks much bite in the judicial context. In the adjudicative context, constitutionally mandated transparency is linked to compliance with procedural and substantive rights of criminal defendants and civil litigants, as well as to resistance to interest-group capture of the courts.³⁰⁶ Judicial data, when disclosed, generally enables the creation of public goods; when held tightly, it typically facilitates state control or private profit. Contrary to the claims of its critics, transparency in this context will not typically “end up shifting the burden of governance outside government.”³⁰⁷ That is, we see no reason to predict that a judicial transparency mandate would necessarily lead to the ills identified when transparency is applied to other government bodies. More specifically, to the extent that this neoliberalism critique has force, it arises from two sources. The first is the possibility that judicial transparency will push litigants to nonpublic forms of mediation or arbitration by analogy to the way that transparency mandates in the administrative state have stifled certain forms of internal consultation.³⁰⁸ The second is the relative stringency of transparency requirements applied to legislative and executive actors compared to the weakness of transparency commands directed at the courts.³⁰⁹ This uneven application of transparency might be glossed as evincing a suspicion of democratic decisionmaking, but not of the more formalist, technical role played by judges. To the extent transparency mandates reflect a neoliberal distrust of the democratic process, this does not require a leveling down: The other option is to impose a high

(“Appealing a FISC surveillance order is virtually impossible in the current system since a defendant might never know such an order had existed in his case or what proof the government had submitted in support of it.”). For an argument that FISC is “breaking the long tradition of public access to judicial opinions, ignoring centuries of legal principles supporting access, and neglecting the bedrock principle of *stare decisis* that American jurisprudence is built upon,” see Thomas Rockhill, Note, *Accessing the Secrets of the Secret Court: The Foreign Intelligence Surveillance Court and the First Amendment Right of Public Access*, 54 U. TOL. L. REV. 469, 476-79 (2023).

305. David E. Pozen, *Transparency's Ideological Drift*, 128 YALE L.J. 100, 145-46 (2018) (quoting Christina Garsten & Monica Lindh de Montoya, *Introduction to TRANSPARENCY IN A NEW GLOBAL ORDER: UNVEILING ORGANIZATIONAL VISIONS* 1, 3 (Christina Garsten & Monica Lindh de Montoya eds., 2008)).
306. For a case in which the interest-group capture of a court was directly linked to the procedural fairness interests of specific litigants, see *Caperton v. A.T. Massey Coal Co.*, 556 U.S. 868, 872 (2009).
307. Pozen, *supra* note 305, at 144.
308. See generally *id.* at 128-29 (summarizing evidence about the effects of open meeting laws).
309. David Pozen hence canvasses transparency mandates directed to Congress and the executive, but not the courts. See *id.* at 123-33, 143-44.

standard of disclosure across the board for government, as well as for other powerful nonpublic actors.

2. Constitutional arguments for opacity

Just as there are elementary forces pressing toward transparency, so too are there pressures for opacity and secrecy. These have foundations in both individual rights and structural principles. They tend to have purchase, however, in specific cases, rather than in general. They are exceptions, not the rules.

On the individual litigant's side, a range of different property and privacy interests can be invoked as warrants for opacity.³¹⁰ Trade secrets are sufficiently akin to property to fall under the shelter of the Takings Clause of the Fifth Amendment.³¹¹ Private persons may have First Amendment interests in privacy in litigation. Sometimes these interests will connect to "intimate privacy."³¹² Other times a constitutional interest in "anonymous speech" might compel a court to derogate from its presumption against anonymous litigation.³¹³ But of course, in many applications such interests will lack a constitutional footing. A court still may recognize nonconstitutional interests

310. See generally Nancy S. Marder, Essay, *From "Practical Obscurity" to Web Disclosure: A New Understanding of Public Information*, 59 SYRACUSE L. REV. 441, 445-47 (2009) (discussing the potential interests in the nondisclosure of court documents). The federal government itself likely lacks a legal interest in judicial data's opacity. The Supreme Court recently affirmed under the "government edicts doctrine" the inapplicability of copyright to the recording of explanatory legal materials created by a legislative body (in that case, state legislative annotations). *Georgia v. Pub. Res. Org, Inc.*, 140 S. Ct. 1498, 1503-04 (2020). The Court had previously applied the same doctrine to hold that "non-binding, explanatory materials are not copyrightable when created by judges." *Id.* at 1504 (citing *Banks v. Manchester*, 128 U.S. 244 (1888)). Extending that logic, it is reasonable to infer that the government does not have an intellectual property interest in judicial data.

311. *Ruckelshaus v. Monsanto Co.*, 467 U.S. 986, 1002-04 (1984). Federal Rule of Civil Procedure 26(c)(G) hence envisages the use of protective orders for "a trade secret or other confidential research, development, or commercial information."

312. See generally Danielle Keats Citron, *Privacy Injunctions*, 71 EMORY L.J. 955, 958 (2022) (describing "intimate privacy" as a function of "the norms that set and fortify the boundaries around intimate life").

313. See *In re Anonymous Online Speakers*, 661 F.3d 1168, 1173-78 (9th Cir. 2011) (recognizing a constitutional protection for anonymous speech, but still permitting mandated disclosure of identities in some instances); Signature Mgmt. Team, LLC v. Doe, 876 F.3d 831, 837 (6th Cir. 2017) (holding that "like the general presumption of open judicial records, there is also a presumption in favor of unmasking anonymous defendants when judgment has been entered for a plaintiff").

in the nondisclosure or nondissemination of information that is embarrassing³¹⁴ or that would inflict emotional harm.³¹⁵

A less-explored set of individual interests sounds in due process. We noted above the potential due process interests in a public hearing, where publicity works in service of impartiality.³¹⁶ But there may be cases where confidentiality is necessary for due process. Transparency might impose such exorbitant reputational costs on the litigant that it deprives them of meaningful access to the courts. This issue is occasionally litigated in the context of request to bring suit under a pseudonym³¹⁷—for instance, when a victim of sexual assault seeks to keep their identity secret. We could imagine a broader argument that court data should be kept secret by default because it is only with secrecy that certain litigants could access the courts safely and securely, and it is impossible to determine in advance who they will be.

In addition to private pleas for secrecy, judges have asserted a relationship between the secrecy and the integrity of their deliberations. Justice Alito, for example, suggested that the leak of a draft abortion opinion “made those of us who were thought to be in the majority in support of overruling *Roe* and *Casey* targets for assassination because it gave people a rational reason to think they could prevent that from happening by killing one of us.”³¹⁸ By analogy to the “deliberative process privilege” asserted by members of the executive branch,³¹⁹ this relationship would justify a secrecy interest in the internal

314. See *Nixon v. Warner Commc'ns*, 435 U.S. 589, 601-02 (1978) (recognizing the risk of “embarrassment” as a cognizable concern with respect to the disclosure of judicial data). An analogy for this interest are the privacy values protected by European law’s “right to be forgotten,” initially recognized in Case C-131/12, *Google Spain SL v. Agencia Española de Protección de Datos*, ECLI:EU:C:2014:317, paras. 20, 89, 91, 99 (May 13, 2014), and now protected under the European Union’s General Data Protection Regulation. Regulation 2016/679, art. 17, 2016 O.J. (L 119) 1.

315. See, e.g., *Nat’l Archives & Recs. Admin. v. Favish*, 541 U.S. 157, 160, 167-68 (2004) (construing a statutory interest in “personal privacy” to include family members’ right to resist disclosure of photos of a deceased relation).

316. See *supra* Part II.B.1.

317. On the manner in which courts treat requests for anonymity in the litigation context, see Eugene Volokh, *The Law of Pseudonymous Litigation*, 73 HASTINGS L.J. 1353, 1367-70 (2022) (identifying a presumption against pseudonymous litigation, and noting its justifications). On the reasons for permitting anonymity, see Lior Jacob Strahilevitz, *Essay, Pseudonymous Litigation*, 77 U. CHI. L. REV. 1239, 1243 (2010) (critically presenting one scholar’s view that “out of concern for those accused of stigmatizing, intentional acts, some courts have allowed them to use pseudonyms or closed the proceedings to the public” (quoting Adam A. Milani, *Doe v. Roe: An Argument for Defendant Anonymity When a Pseudonymous Plaintiff Alleges a Stigmatizing Intentional Tort*, 41 WAYNE L. REV. 1659, 1698 (1995)).

318. Liptak, *supra* note 11.

319. The Court has described the deliberative process privilege as “a form of executive privilege” that “protect[s] agencies from being ‘forced to operate in a fishbowl.’” U.S. *footnote continued on next page*

discussions and processes by which legal and factual issues are resolved within the courthouse. Although we are not aware of binding authority on this point, courts' willingness to recognize this interest when invoked under an Article II rubric strongly suggests they would be open to analogous arguments running under an Article III label. This analogy would also explain why neither the CM/ECF nor the PACER system contain any deliberative documents. For this reason, a federal statute tracking the Federal Advisory Committee Act³²⁰ and mandating the disclosure of all judicially produced materials would unconstitutionally impinge on the inherent Article III authority to deliberate that is antecedent and adjunct to the Article III judgment power.

A more controversial extension of the same secrecy argument might encompass both aggregated court records and summary statistics for individual judges. As discussed in Part I, reports produced by the judiciary omit judge-specific information out of a concern that it will be "taken out of context to be misinterpreted."³²¹ This worry is pitched at a high level of generality, so it is hard to discern exactly what is at stake. One version of this concern might be as follows: Imagine that judge-specific information related those greater sentencing disparities by race or gender on the part of *Judge A* than on the part of *Judge B*. Lacking any realistic (or ethical) ways for the judge to explain their practice across cases to the general public,³²² the worry is that the release of this information might have one of two undesirable effects. *Judge A* might be tempted to increase their punitiveness with one group or *Judge B* might be tempted to relax the attention paid to another group's cases. Alternatively, the release of judge-specific information might kindle a sentencing judge's consciousness of race or gender in constitutionally problematic ways insofar as the judge's decision changes simply because an awareness of a defendant's race (holding all else constant).³²³

The force of this last argument is far from clear. As one federal district court judge has candidly observed, the risk that data "may be misinterpreted

Fish & Wildlife Serv. v. Sierra Club, Inc., 141 S. Ct. 777, 785 (2021) (quoting EPA v. Mink, 410 U.S. 73, 87 (1973)). It "shields from discovery 'advisory opinions, recommendations and deliberations comprising part of a process by which governmental decisions and policies are formulated.'" *Id.* (quoting NLRB v. Sears, Roebuck & Co., 421 U.S. 132, 150 (1975)).

320. 5 U.S.C. §§ 1001-1014.

321. ADMIN. OFF. OF THE U.S. CTS., *supra* note 175, at 21-22.

322. Interview with Source No. 3, *supra* note 83.

323. The Supreme Court has recently narrowed the constitutionally permissible scope for race-conscious decisionmaking. *See* Students for Fair Admissions, Inc. v. President & Fellows of Harvard Coll., 143 S. Ct. 2141, 2166 (2023). By pointing to that decision and its colorblind rule, we do not mean to imply agreement with its reasoning or conclusion. Nor do we mean to foreclose arguments that consideration of race in the judicial context might rest on different legal grounds.

and judges may be criticized is not a valid reason to withhold information.”³²⁴ Judges have tenure and salary protections precisely so as to insulate them from the persuasive force of public criticism.³²⁵ But the argument for keeping aggregate data and summary statistics from the general public assumes that judges are so psychologically fragile that public criticism will cause them to change their behavior. If that assumption were true, however, it would justify far more extensive prohibitions on speech critical of the federal judiciary. Obviously, that is not part of our tradition.³²⁶

On the other hand, it is not clear that judges should always avoid changing their behavior based on new learning that has been derived from judicial data. To illustrate this, imagine a judge who is apprised of the fact that they consistently punish men more harshly than similarly situated women. This pattern may elude the judge because of the hectic pace of their courtroom. Once it is drawn to their attention, however, they might reflect on the way that their preconceptions of gender shape their sentencing practices—and then engage in valuable behavioral change. Consider too the same hypothetical, albeit centered around race rather than gender: Would the increasing constitutional skepticism of any and all race consciousness mean that such self-correction was impermissible?

To summarize, it is implausible to assume that judges are “snowflakes” who will melt at the first hint of criticism. We should also not assume that judges should be inured from data-driven criticism in all cases. In our view, the judiciary’s constitutional arguments for secrecy are on weak constitutional footing and fail to persuade as a matter of public policy.

III. Judicial Data as Public Good: An Argument and a Legislative Proposal

Judicial data, then, is a rich and complex world. It could (and to some extent does) generate concentrated private profits or improve public services

324. T.S. Ellis, III, *Sealing, Judicial Transparency and Judicial Independence*, 53 VILL. L. REV. 939, 939, 942 n.6 (2008).

325. On the public-facing threat to judicial independence, see John A. Ferejohn & Larry D. Kramer, *Independent Judges, Dependent Judiciary: Institutionalizing Judicial Restraint*, 77 N.Y.U. L. REV. 962, 969 (2002) (“Efforts to overpower the judiciary are as likely to emanate from the people themselves, acting through faithful representatives, as they are from defects in political agency. This concern is, if anything, more pronounced today than it was in Madison’s time given the development of political parties, organized lobbies, and the like.”).

326. That said, there is a tradition of “lawyers [who] were disciplined for exceeding the bounds of decorum and responsibility in their attacks on judges.” Hal R. Lieberman, *Essay, Should Lawyers Be Free to Publicly Excortiate Judges?*, 25 HOFSTRA L. REV. 785, 788-95 (1997) (documenting examples from state and federal courts).

for all. Some such data is captured, some is lost, and much is never made available to the public at large—all with scant guidance from the Congress that has primary power and responsibility.

This is, to say the least, a lost opportunity. Using its broad Necessary and Proper power, Congress *ought* to direct with more precision how judicial data is captured and used—especially given courts’ failure to use their inherent powers to do so.³²⁷ This Part offers a normative argument for new legislation organized around the vision of judicial data as a public good.³²⁸ We think Congress should enact legislation designed to advance the public goods such as more equitable and accessible adjudication that can be derived from such data. This means enabling broad access to data, with exceptions identified in Part II, via interfaces that empower a wide array of potential users (including individuals, firms, and academics) to tap that resource. While data conglomerates such as RELX and Thomson Reuters have a role to play in new data economies, new legislation must be carefully designed to minimize information oligopolies that maximize private profit to the detriment of public goods. Emphasis on private gains, indeed, conduces now to “mediocre content and incorrect predictions” thanks to the lack of competitive pressures.³²⁹ To ensure that the large benefits to be derived from judicial data are widely shared, new legislation is needed.

We develop this normative argument in two steps. First, we make a general case for treating judicial data as a public asset to be cultivated and disseminated for the public good. Second, we offer a series of specific interventions that a new law could pursue to advance this vision of judicial data. Of course, Congress is famously slow to act. Nevertheless, we think this is an opportune time to reconsider the statutory regulation of judicial data. In Congress, multiple bills have already been proposed to reduce or eliminate fees for court records.³³⁰ Leading scholars have called for open access under the banner “Free PACER,”³³¹ or demanded that Congress or the courts “Tear Down

327. See Alexandra D. Lahav & Peter Siegelman, *The Curious Incident of the Falling Win Rate: Individual vs System-Level Justification and the Rule of Law*, 52 UC DAVIS L. REV. 1371, 1375 (2019) (finding large changes in civil litigation and charging that “[t]he courts have largely ignored system-level changes”).

328. For a general discussion of public goods, see John O. Ledyard, *Public Goods: A Survey of Experimental Research*, in THE HANDBOOK OF EXPERIMENTAL ECONOMICS 111, 111-13 (John H. Kagel & Alvin E. Roth eds., 1995) (defining public goods in terms of the empirical phenomenon of freeriding).

329. LAMDAN, *supra* note 37, at 131.

330. See, e.g., Open Courts Act of 2021, S. 2614, 117th Cong. (2022); Open Courts Act of 2021, H.R. 5844, 117th Cong. (2021); Open Courts Act of 2020, H.R. 8235, 116th Cong. (2020); Electronic Court Records Reform Act of 2019, S. 2064, 116th Cong. (2019).

331. See Gelbach, *supra* note 46, at 328-29.

This Judicial Paywall.”³³² Although efforts to increase access deserve applause, these rallying cries speak only to the tip of the information iceberg. Many other difficult choices need to be made about data access, and numerous other policy determinations are needed to govern the production, collection, and management of data *before* disclosure—i.e., before the point at which “free PACER” has bite. By mapping a range of other consequential decisions upon which the value and meaning of “free PACER” depend, we hope to wrestle the public debate out of its myopic rut.

A. Judicial Data as a Public Asset

The broad congressional authority over judicial data and the absence to date of sustained and general legislative action invite the normative questions of what sorts of policy *should* govern federal court data. We think the answer to the normative questions flows from a recognition of judicial data as a public asset that should be mined and distributed for the good of the public at large.

At this point, we think it is clear that judicial data is a valuable asset. Across the economy as a whole, data is recognized with growing frequency as an increasingly important “raw material of business” and “a new form of economic value.”³³³ Its exploitation yields “rents” that can be allocated in more or less regressive ways.³³⁴ Decisions as to when and who can transform it into an asset determine how large profits are and where they are allocated.³³⁵ Today, firms routinely engage in sweeping collection of data regardless of a firm’s “imaginative reach or analytic grasp” in the hope such data will “eventually be useful.”³³⁶ Judicial data, like other “raw materials of business,” is an asset that can be exploited, with surplus value being shared in more or less regressive ways.

332. Sarath Sanga & David Schwartz, Opinion, *Tear Down This Judicial Paywall*, WALL ST. J. (Dec. 13, 2020, 6:00 PM ET), <https://perma.cc/NP87-72JD>.

333. VIKTOR MAYER-SCHÖNBERGER & KENNETH CUKIER, *BIG DATA: A REVOLUTION THAT WILL TRANSFORM HOW WE LIVE, WORK AND THINK* 5 (2013). For a broader argument about public control of personal data, see Aziz Z. Huq, *The Public Trust in Data*, 110 GEO. L.J. 333, 340-43 (2021), which describes the history of data economies.

334. See Kean Birch, Margaret Chiappetta & Anna Artyushina, *The Problem of Innovation in Technoscientific Capitalism: Data Rentiership and the Policy Implications of Turning Personal Digital Data into a Private Asset*, 41 POL’Y. STUD. 468, 475, 480 (2020).

335. See, e.g., Thomas Beauvisage & Kevin Mellet, *Datassets: Assetizing and Marketizing Personal Data*, in *ASSETIZATION: TURNING THINGS INTO ASSETS IN TECHNOSCIENTIFIC CAPITALISM* 75 (Kean Birch & Fabian Muniesa eds., 2020) (describing one proposal to give individuals high control over their data which would lead to widely distributed value creation).

336. Marion Fourcade & Kieran Healy, *Seeing Like a Market*, 15 SOCIO-ECON. REV. 9, 13 (2017).

The case for treating judicial data as a *public* asset starts with its origin. All data in the economy is, in a sense, the product of social activity and common labor.³³⁷ This claim is all the more forceful in the context of judicial data, as opposed to data generated through transactions in the market at large. As Part I explained, judicial data is the result of both dispersed litigants' work and also the publicly funded actions of Article III officials. It is not the product of private enterprise. It is rather coproduced by the public and by officials funded by the public purse. As Judith Resnik writes,

[N]othing is casual about the public's relationship with courts. Courts are funded by the public fisc and staffed by government employees working in buildings owned or rented by the government and subject to a thicket of government-crafted regulation. Courts are also public in the sense that the members of the public are entitled to file cases, to watch proceedings involving others in courts, and to know the identity of judges and staff, their salaries (set by law), their budgets, and the rationales for their rulings.³³⁸

Judicial data, as the product of these public actions, should be used for the broad benefit of the public.

In addition, the assumption that judicial data is a public asset is almost baked into the system. There are innumerable ways that a dispute between two people can be resolved. Each comes with its own costs, benefits, and norms. Resolving disputes in court is one of the most public ways that disputes can be resolved. It is a public-facing forum in the sense of being open by default to the public, and being shaped and channeled directly though (public) law. It also receives more public subsidies than the immediate substitutes, such as arbitration or mediation.³³⁹ Unlike these alternatives to official courts, litigation also generates generally binding precipitates—binding judgments and even new rules of law. The thoroughly public character of the judicial process is part of what makes it distinctive. Publicity communicates content to the community, disciplines the truthfulness of parties, and serves as a justification for the making of new law in the form of precedent.³⁴⁰ Publicity

337. For a similar argument applied to personal data generally, see Huq, note 333 above, at 338 (“[W]e should view this [personal] data not as an aspect of individual action or a particular firm’s ingenuity, but as a shared asset—one realized through the entangled social interactions of the many, and one capable of vindication only by recognizing the many rather than the one as pivotal.”).

338. Judith Resnik, Feature, *Diffusing Disputes: The Public in the Private of Arbitration, the Private in Courts, and the Erasure of Rights*, 124 YALE L.J. 2804, 2835 (2015).

339. See Charles W. Tyler, Note, *Lawmaking in the Shadow of the Bargain: Contract Procedure as a Second-Best Alternative to Mandatory Arbitration*, 122 YALE L.J. 1560, 1585 (2013) (noting that “arbitrators . . . financially depend on the patronage of repeat players” and “legislatures have not provided public funding of arbitration”).

340. Of course, there has long been a lively debate about “the legitimacy of even very conventional forms of federal common law.” Larry Kramer, Essay, *The Lawmaking Power of the Federal Courts*, 12 PACE L. REV. 263, 264 (1992).

may be part of the appeal to litigants, and at a minimum, it must be an assumption of everyone passing through the courthouse doors. In short, public access to information is *inherent* in the judicial process.

The case for judicial data as a public asset is also instrumental. Judicial data is a public asset because it can be used to improve the quality of governance and can help those interested in understanding how they are being governed. It can illuminate the operation of one branch of our national government. Constitutional lapses can be brought to light by examining specific items of whole tranches of judicial data.³⁴¹ Such data, if made available as an information commons, can serve as a quintessential public good: It can work as a resource for researchers and reformers aiming at a better understanding, or an improvement in, judicial institutions.³⁴² And it can achieve these goals without intruding on other valuable uses. That is, judicial data is both nonrivalrous and nonexhaustible in the classic sense of a public good.³⁴³

Private firms are unlikely to realize this public-good quality of judicial data. Not least because they have no pecuniary interest in trying to realize gains that accrue widely to the public at large beyond their paying clients.³⁴⁴ Nor are they concerned with fostering free inquiry into the operation of our government.³⁴⁵ Instead, their regulation of access to information for the sake of profit is likely to hinder the public's "ability to seek information without barriers to access"³⁴⁶ that is at the heart of democratic life. The need for "more democratic, horizontal forms of economic coordination"³⁴⁷ when it comes to judicial data is therefore functionally compelling.

341. A good example is the range of issues identified by the empirical study of the "Six-Month List." See de Figueiredo et al., *supra* note 171, at 365-72 (listing normative and legal concerns); see also *infra* notes 395-99 and accompanying text.

342. See Pah et al., *supra* note 21, at 134-36.

343. To be sure, the sharing of this data may harm some actors by taking away their opportunities at rent seeking—but that is, or should be, part of the point. Cf. J.H. Reichman & Pamela Samuelson, *Intellectual Property Rights In Data?*, 50 VAND. L. REV. 51, 155-61 (1997) (identifying public-good aspects of data, and arguing that the law should actively elicit them).

344. *But see* Fourcade & Healy, *supra* note 336, at 13 (describing commercial incentives to gather data, even when there is no immediately evident use).

345. For an example of how firms work around legal constraints on the use of data to maximize profits, see Brenda M. Simon & Ted Sichelman, *Data-Generating Patents*, 111 N.W. U. L. REV. 377, 380 (2017) ("Google's ability to capture information about its users, and to continue using this proprietary information after its patents expire or are invalidated, provides it the ability not only to improve its search functionality, but also to leverage its proprietary data to create a superior product in the distinct market of targeted online advertising.").

346. LAMDAN, *supra* note 37, at 128.

347. Sanjukta Paul, *Antitrust as Allocator of Coordination Rights*, 67 UCLA L. REV. 378, 430 (2020).

That said, suggesting that court data is a public asset does not mean there is no role for private actors.³⁴⁸ As described above, much court data is produced exogenously, and so there is no escaping the role of private actors in data production and capture. In addition, the management of any public asset can involve private actors. The government can contract with private firms to manage public assets, or it can make those assets available subject to regulation that has the effect of making public use of private interventions. The presently privileged role played by data cartels with respect to judicial data is an example of this possibility.³⁴⁹ All of this is to say that calling judicial data a public asset tells us that it should be produced and used in ways that keep the public good in mind. It need not be a wholly state-owned enterprise. The government should hence act as a comprehensive regulator, and not necessarily as an owner and operator, when it comes to judicial data.

B. Reforming the Stewardship of Judicial Data

The possibility of asking Congress to regulate the data of the federal courts as a public asset of significant untapped value opens up a number of unresolved policy questions and potential points of reforms at all stages of the information flow. Mapping out the proper stewardship of judicial data, we take up first the issue of data production and capture, followed by a separate discussion of data access. As distinct from the arguments focused solely on the greater disclosure of judicial data for academic study, our central aim here is to identify and explore ways of enhancing the creation of broadly beneficial public goods from judicial data.

1. Data production and capture

To begin with, there are difficult questions at the margins as to what information should be produced and captured in the first instance. Debates around the collection of surveillance information, for example, usefully flag arguments why sometimes it is better to not capture information rather than to regulate its use.³⁵⁰ Some information may be so high risk and offer such low public value that it ought not be collected in the first place. Social security numbers or other personally identifying information, or personal details of

348. See Huq, *supra* note 333, at 336-37 (observing that a “[public] asset need not be in the physical custody of the state: private actors can maintain day-to-day control, while the public remains the beneficial owner”).

349. See *supra* Part I.C.3.

350. See, e.g., PRIV. & C.L. OVERSIGHT BD., REPORT ON THE TELEPHONE RECORDS PROGRAM CONDUCTED UNDER SECTION 215 OF THE USA PATRIOT ACT AND ON THE OPERATIONS OF THE FOREIGN INTELLIGENCE SURVEILLANCE COURT 159-61 (2014), <https://perma.cc/LH8M-NJFF>.

plaintiffs in cases involving violations of intimate privacy, are potential examples of data that could be ex ante regulated through prohibitions on disclosure or rules for secure handling.³⁵¹ But for the vast majority of judicial data, no fundamental values are endangered by the mere fact of collection.³⁵² Congress should endorse, then, a default norm of collection, subject to exceptions that serve these constitutionally inflected values.

Merely endorsing the desirability of better data collection and production does not make it happen. Our positive account in Part I reveals how data production and collection can be improved in terms of accuracy and consistency. It seems plain to us that once a decision is made to produce and collect data, accuracy and consistency should at least be among the relevant policy goals. Our account suggests that the federal courts are falling short on both counts. Congress, therefore, should step in to improve both accuracy and consistency. In this section, we draw on the descriptive work of Part I to develop two examples of fruitful legislative intervention at the threshold moment when judicial data is produced.

First, consider the value of increased legislative control as to whether data production and capture should be endogenous or exogenous. Our suggestion that court data is a public asset does not directly answer this question. Instead, it suggests that answers to such queries should depend on what mode of data production is likely to maximize public goods. The choice between these options, therefore, should depend on the incentives and capacities of the possible public and private actors.

The current approach, as described above, relies heavily on exogenously produced data. This approach makes sense for the kind of data that is useful for the resolution of the case. For better or worse, the adversarial system generally puts the onus on parties and their attorneys to build a record and frame issues for resolution.³⁵³ Consequently, one party or another will have incentives to produce and capture exogenous data that is useful for resolution of the case in their favor—even without legal requirements. Yet even this assumption breaks down when parties have an interest in coordinating or colluding to reach a result that is optimal for them, but that has externalities. For example, parties

351. Interview with Source No. 6, *supra* note 99. Currently, parties are responsible for redacting social security numbers, and PACER allows searches using full social security numbers for debtors. See *FAQs: Case Management/Electronic Case Files (CM/ECF)*, *supra* note 77 (“At login to CM/ECF, a message reminds attorneys of their responsibility to redact this private information from the documents they file.”); *Are Full Social Security Numbers Displayed on Court Documents?*, PACER, <https://perma.cc/7CVY-8UGG> (archived Apr. 17, 2024) (“[A] full SSN can be used to perform searches to identify debtors.”).

352. See generally *supra* Parts I.A-B.1 (documenting the kinds of information collected).

353. See David Alan Sklansky, *Anti-Inquisitorialism*, 122 HARV. L. REV. 1634, 1636 (2009) (“Avoiding inquisitorialism is taken to be a core commitment of our legal heritage.”).

may agree to share discovery information among themselves and the court, while shielding it from public scrutiny.

The assumption of adequate private incentives to produce is even less plausible when it comes to the kind of court data that might be useful for broader analyses but is of little relevance to winning the case at hand. For example, parties do not directly gain by producing accurate NOS codes because they do not influence how a case is resolved. The NOS code has no apparent relevance to the individual adjudication. A party is no more likely to prevail if they code their case correctly or not. There is hence little incentive for a party to properly assign a NOS code.³⁵⁴ But NOS codes are quite relevant outside of adjudication for scholars and policymakers seeking to study large quantities of court data. The federal courts rely on NOS codes when trying to determine the workload of individual courts, which informs the judiciary's recommendations for the creation of additional judgeships.³⁵⁵ NOS codes are (or should be) useful for scholars and policymakers seeking to study large quantities of court data because they are (or should be) an indicator of case type that allows researchers to ask all sorts of important questions.³⁵⁶ In effect, NOS codes are a public good that private parties will likely underproduce. It hence may be appropriate to lean more on endogenous data production of NOS codes because litigants are likely already producing all of the data that aligns with their incentives. We think also this is more likely to be effective than a regime that requires private parties to produce such data accurately, or else face sanctions. Further, we think that sanctions in private production regimes are unlikely to be timely enforced given the other demands on judges' time and attention.

As Part II demonstrated, Congress has broad latitude to demand endogenous document production, except perhaps where doing so impinges on an Article III claim to inherent authority.³⁵⁷ Beyond a legislative demand that judges record and publish in-chambers deliberations or early drafts, it is hard to see why such a concern would arise in respect to data such as NOS codes. This is not to suggest that Congress should maximize the compelled demand of endogenous data. One could imagine stiff resistance from the judiciary if Congress were to, for example, require that dispositive decisions always be in

354. See *supra* note 60-62 and accompanying text.

355. See Judith Resnik, *Trial as Error, Jurisdiction as Injury: Transforming the Meaning of Article III*, 113 HARV. L. REV. 924, 983-84 (2000) (discussing "case weights" that, in turn, depend on the coding of case types); PATRICIA LOMBARD & CAROL KRAFKA, FED. JUD. CTR., 2003-2004 DISTRICT COURT CASE-WEIGHTING STUDY: FINAL REPORT ON THE SUBCOMMITTEE ON JUDICIAL STATISTICS OF THE COMMITTEE ON JUDICIAL RESOURCES OF THE JUDICIAL CONFERENCE OF THE UNITED STATES 1-2 (2005), <https://perma.cc/HB2K-JXR3> ("Civil case types were based on the set of nature-of-suit codes . . .").

356. For examples, see Part III.C below.

357. See *supra* Part II.A.

writing.³⁵⁸ Although it is important to see that such resistance would have no firm foundation in Article III, such a broad mandate would nevertheless be unnecessary and counterproductive. There is also no Article III obstacle to legislation requiring that parties' motions or pleadings be in writing, or that transcripts of certain proceedings be preserved and made available.³⁵⁹

Second, a major source of data loss is the inefficiency and variance of front-end data capture. Now, data is routinely captured in forms that are difficult to access subsequently (for example, because of an absence of searchability) or in ways that vary across courts (or even across chambers within the same courthouse).³⁶⁰ Especially where data is endogenously produced, as part of the necessary and proper stewardship of federal court data, Congress should call upon federal courts to capture data in ways that promote broad public accessibility and uniform availability. While judges might well bristle at a statute requiring a written opinion to be published in response to every motion for summary judgment (not unlike a new Texas law³⁶¹), we think there would be (or at least should be) less resistance to a statute requiring that every published opinion on summary judgment be labeled with the digital tag "Summary Judgment." Or more generally, we hope there would be little resistance to a statute directing the courts to author and publish a centralized "code book" for use by all the federal courts. This standardization should come at little financial cost to courts, and we are not aware of any strong arguments why the courts or public benefit from disparate coding practices (other than the resistance to change and the costs of learning new approaches). And doctrinally, once a decision is made to produce and capture certain judicial data, efforts to standardize that process should implicate even weaker arguments from inherent Article III authority.

Moving data production from exogenous to endogenous, and from individuated to standardized, sounds like a call for more central bureaucracy. But just because the government has an interest in—if not a public responsibility for—capturing data, that does not mean the government needs to internalize this work. Privatization, properly designed, is still an option.

358. Similar issues may arise if Congress attempts to regulate the Supreme Court's use of the so-called Shadow Docket. See William Baude, *Foreword: The Supreme Court's Shadow Docket*, 9 N.Y.U. J.L. & LIBERTY 1, 30-31 (2015) (noting the absence of clear rules guiding such practice without asking about the utility of new law on the matter).

359. See *supra* Part II.A.2.

360. See *supra* Part I.D; see also *supra* notes 90-93 and accompanying text (discussing the variability of data capture between courts); *supra* notes 131-43 and accompanying text (discussing searchability).

361. See Act of May 24, 2023, 2023 Tex. Sess. Law Serv. ch. 209 (West) (codified at TEX. CIV. PRAC. & REM. CODE ANN. § 51.014(g) (West 2023)) (requiring a written opinion when an appellate court declines to hear a certified interlocutory appeal).

Privatization typically involves either “removing certain responsibilities, activities, or assets from the collective realm” or “retaining collective financing but delegating delivery to the private sector.”³⁶² Debates about privatization are long running. There are long lists of well-understood costs and benefits, but little agreement.³⁶³ Our thick description in Part I suggests that at least some aspects of endogenous data production and capture are amenable to privatization. Privatization would be most valuable when there is agreement that data production and capture are appropriate, but the incentives for parties and courts to act are weak. For example, with regard to front-end data, Congress might direct a private firm to develop a predictive algorithm to assign NOS codes based on a scan of every machine-readable complaint.³⁶⁴ Here, privatization might be instrumentally valuable because private firms often have stronger incentives than parties to standardize data and ensure it is error free. In part, this is because a firm can be tasked with a specific responsibility and then monitored to evaluate its performance. Courts and parties, by contrast, lack incentives to ensure that filings in one case look similar to filings in another. Their data-related practices are more dispersed than those of a specific firm, and hence harder to monitor. Nor are they likely worried about data-entry errors that do not affect the disposition because these errors only generate costs external to the litigation. Privatizing these functions may align incentives better, and at lower cost.

Doctrinally, there is no inherent Article III power argument of the sort canvassed in Part II that would impede many forms of privatization.³⁶⁵ Rather, Congress has wide discretion to ascertain when private actors should manage or supplant state action respecting judicial data flows. The lack of constitutional constraints also implies that broad-brush arguments against privatization should be questioned.³⁶⁶ But recognizing judicial data as a public asset means that any efforts at privatization should be public regarding. That means ensuring that private firms are not capturing *all* of the value created by judicial data. It also means ensuring that they are managing that data in ways that benefit the public, even if they would not be economically efficient for the firm.

362. JOHN D. DONAHUE, *THE PRIVATIZATION DECISION: PUBLIC ENDS, PRIVATE MEANS* 215 (1989).

363. For a useful overview, see Jon D. Michaels, *Privatization's Pretensions*, 77 U. CHI. L. REV. 717, 717-18 (2010) (crisply summarizing both a case “for” and a case “against” privatization).

364. This is yet another argument for moving certain data-production duties from exogenous to endogenous.

365. See *supra* Part II.A.2.

366. For an example of a very broad argument against privatization that is also deeply learned and sophisticated, see CHIARA CORDELLI, *THE PRIVATIZED STATE* 7 (2020).

2. The public disclosure of judicial data

Once data is produced and captured, the question arises whether the general public will have access, and if so, under what terms.³⁶⁷ This is the terrain on which most debates have occurred, albeit typically focused on PACER's price. Our study of the federal courts' information ecosystem, however, reveals a wider array of questions, and a broader array of ways in which judicial data disclosure can be regulated.

Our starting point is a presumptive norm favoring the transparency of governmental operations. As Part II demonstrated, there are powerful, a priori constitutional grounds for transparency—both from individual rights and structural perspectives.³⁶⁸ Public assets such as judicial data, which is the fruit of public investment in the courts, should be presumptively available to the public. Thus Congress should endorse a background principle of transparency of judicial action.

But, of course, for legal and policy reasons, the norm of transparency is subject to exceptions. These exceptions might be necessary to protect individual rights or structural values. Here the metaphor of walls and hurdles helps explain how Congress might craft exceptions to the transparency norm.

First, walls are justified where the goal is to prevent access to specific pieces of data altogether. This might be for due-process-like reasons.³⁶⁹ A victim of intimate abuse, for instance, may not have meaningful access to the federal courts to seek injunctive relief against a website disseminating their intimate photos without a wall around personal identifying information. Sensitivity to these concerns is important in making data policy choices because ex ante choices about transparency can have ex post effects on judicial access. For example, a categorical rule against anonymity in litigation would likely deter some plaintiffs from filing suit.³⁷⁰

There also might be structural reasons for setting up walls. For example, inherent Article III power may compel walls around the deliberative portion of the judicial process.³⁷¹ In crafting “structural” exceptions to transparency defaults, the federal law of evidentiary privileges offers lessons. Federal agencies are subject to disclosure requirements but can protect certain

367. There may also be questions about when and how information can be shared within the judiciary—for example, between chambers and a clerk's office—or between different courts respecting pending cases raising the same legal question.

368. *See supra* Part II.B.

369. *See supra* Part II.B.1.

370. For a cogent discussion of the use of pseudonyms in the context of revenge porn cases, see Strahilevitz, note 317 above, at 1240-47.

371. *See supra* Part II.B.2.

internal information under the deliberative process privileges.³⁷² Lawyers and clients can be asked to testify about any relevant information. But their confidential communications for the purpose of seeking legal advice are protected by the attorney-client privilege and work-product doctrine.³⁷³ Both privileges are designed to further specific policy values—to allow agencies to make good decisions and to allow lawyers to give good advice on the basis of complete information. Similarly, a judicial deliberations privilege might presumptively shield internal court materials from disclosure. If litigants request their disclosure, they would have to present powerful, specific arguments in justification.³⁷⁴ Judges should be able to circulate drafts of opinions confidentially on the theory that this improves the quality of their work product.

Implicit here is a further question as to whether walls should be narrowly or broadly drawn. A judicial-deliberations privilege would be a wholesale, categorical rule. So too would be a policy of redacting social security numbers. Congress has the capacity to take on such wholesale wall building. An added benefit is that wholesale walling lends itself to automation, which in turn aspires to lower cost and greater accuracy.³⁷⁵

Nevertheless, some retail wall building will be needed. Congress might intervene in retail wall building not by attempting the impossible task of identifying *ex ante* situations in which nondisclosure is warranted, but instead by reforming the process around which such decisions are made. In particular, these retail wall construction situations are prone to abuse because there is no adversarial process at work.³⁷⁶ In other areas, the law has developed at least two

372. *See, e.g.*, *U.S. Fish & Wildlife Serv. v. Sierra Club, Inc.*, 141 S. Ct. 777, 786-88 (2021) (applying the deliberative privilege doctrine); *see also In re Sealed Case*, 121 F.3d 729, 737 (D.C. Cir. 1997) (describing the doctrine).

373. *See, e.g.*, *Upjohn Co. v. United States*, 449 U.S. 383, 389 (1981) (discussing the sources and scope of the privilege under federal law); *Hickman v. Taylor*, 329 U.S. 495, 510-11 (1947) (discussing the work-product doctrine); *see also* David Freeman Engstrom & Jonah B. Gelbach, *Legal Tech, Civil Procedure, and the Future of Adversarialism*, 169 U.P.A. L. REV. 1001, 1077-1080 (2021) (discussing the work-product doctrine in the context of legal technology).

374. To be clear, we do not anticipate a regime of private rights of action to force the disclosure of judicial information.

375. On the potential for cost savings from automation in the legal industry, see Gillian Scott, *What is Legal Automation? How Law Firms Use AI to Increase Efficiency and Add Value for Clients*, *LEXPRT* (Nov. 1, 2021), <https://perma.cc/PCC7-DXSP>. That said, advances in machine learning and the application of LLMs may make the application of tailored privilege rules easier over time.

376. For example, as we noted above, a party may have reasons to produce data for judicial examination but also to agree (or collude) with its litigation opponents to shield that information from the public. Indeed, Seth Endo has found that parties with a shared interest in confidentiality may file joint requests to file materials under seal, not

footnote continued on next page

strategies for dealing with such a dearth of adversity. First, the law sometimes creates opportunities and subsidies for intervention by interested parties.³⁷⁷ In the context of data production, transparency-forcing interventions of this sort might be effective in highly salient or commercially profitable contexts. In the majority of cases, however, we suspect there will be no plausible intervenor at hand. That brings us to the law's other approach: creating adversity by creating an adversarial office. In some cases, judges will appoint counsel to represent a position without a party.³⁷⁸ Courts should be encouraged to take up such tools to address incentive deficits on their own initiative.

A second form of data-access regulation is what we call the "hurdle." Congress, too, should attend to the hurdles around federal court data. The most obvious hurdle, and the one to which Congress has paid the closest attention, is fees. PACER charges a fee per page.³⁷⁹ This fee-per-page model of course favors those with resources. So replacing a wall with a monetary hurdle increases transparency—but only for those who can afford it. The current PACER fee model also favors those interested in small numbers of individual documents (for example, the reporter seeking details of a particular case of police misconduct or the celebrity stalker) over those interested in large quantities of information (for example, the scholar seeking to study rates of police misconduct across time and place or the profit-minded debt-collection agency).³⁸⁰ One thorough estimate found that for scholars to download all district court documents for civil and criminal cases filed in just one year (2016), the cost would have been more than \$5 million.³⁸¹

infrequently reciting an erroneously low standard to justify the order. *See* Seth Katsuya Endo, *Contracting for Confidential Discovery*, 53 U.C. DAVIS L. REV. 1249, 1252-54 (2020).

377. When lawyers for plaintiffs and defendants propose a class-action settlement that might be collusive, for example, the law allows objectors to raise concerns with the district court, mitigating information gaps created by the lack of adversity. *See* FED. R. CIV. P. 23(e)(5).

378. *See, e.g.,* Katherine Shaw, Essay, *Friends of the Court: Evaluating the Supreme Court's Amicus Invitations*, 101 CORNELL L. REV. 1533, 1535 (2016) ("Amicus invitations . . . which generally arise when one party to a case declines either to participate at all, or to take a particular position, before the Court . . . come about once each Term."). For another example, under the USA FREEDOM Act of 2015, the Foreign Intelligence Surveillance Court must appoint one of the designated amici curiae when a case raises an important issue of law on an issue lacking adversarial presentation; a court may also see fit to appoint an amicus in other cases at the court's discretion. Pub. L. No. 114-23, § 401, 129 Stat. 268, 279 (2015) (codified as amended at 50 U.S.C. § 1803(i)(2)).

379. *See* PACER CASE LOCATOR, *supra* note 153.

380. *See supra* Part I.C.1.

381. *Modelling PACER Costs: A Technical Review*, SCALES (Dec. 21, 2020), <https://perma.cc/RF3N-N762> ("From the user side, it is prohibitively expensive to compile large amounts of case data. We estimated it would take somewhere between \$5.5 million and \$5.75 million to access and download all case data for Civil and Criminal cases filed in 2016 in the 94 district courts.").

But PACER's model is not the only imaginable model for regulating docket access. Users could instead pay a single fee for as many documents as they want. This regime, which is indifferent to the volume of data demanded, would favor users seeking larger numbers of documents (who would pay a lower marginal cost per document)—and likely disincentivize the small-volume user at the margin. Yet another approach would charge market prices for data, charging more for easily monetizable data (for example, many complex commercial cases) and less for cases with less commercial use (for example, likely the majority of criminal cases). Hurdles can further be adjusted with subsidies. A subsidy could be available on the basis of need, pushing back on any skew in favor of the well resourced.³⁸² A subsidy could also be made available for certain uses, such as those related to legal education and research.³⁸³ Or a subsidy could be given to users of certain qualities or types of data.³⁸⁴ For example, data about criminal adjudication might be available at a discount to encourage the study of data that otherwise would not be exploited for profit. The law already contains examples of how this might be done. The doctrine of fair use in copyright law is an example of a subsidy available to both commercial and nonprofit actors, but only for specific ends.³⁸⁵

In addition to monetary hurdles, data accessibility can be undermined by another kind of hurdle: data format.³⁸⁶ An individual source might be more or less accessible depending on how it is memorialized. The ease of locating relevant individual sources depends on the searchability of the data.³⁸⁷ Think of the relative ease of finding certain statements in an audio recording as opposed to in a transcript. Or contrast how easy it is to search a text file in Microsoft Word versus an image-only PDF document created by Abode Acrobat. Recall too that Part I identified other examples of design choices that limit access in one form or another: Most obviously, judge-specific data is available only for those willing to assiduously pick through PACER docket

382. Notably, the grant of a petition to proceed in forma pauperis does not automatically exempt users from PACER fees. See *Pricing FAQs*, PACER, <https://perma.cc/JM4R-S8E7> (archived Apr. 17, 2024).

383. Currently, researchers may apply for PACER fee waivers. See *Fee Exemption Request for Researchers*, PACER, <https://perma.cc/FRW8-FD3Z> (archived Apr. 17, 2024).

384. Currently, PACER exempts the first \$30 of fees per quarter and limits fees for a single document to \$3. See *Frequently Asked Questions: Pricing FAQs*, *supra* note 382.

385. In general, judges have glossed fair use in broad terms. See Barton Beebe, *An Empirical Study of U.S. Copyright Fair Use Opinions, 1978-2005*, 156 U. PA. L. REV. 549, 620-22 (2008) (describing judicial practice as deeply “flawed”).

386. Data access challenges could be thought of as a form of “queueing,” an alternative to pricing. See generally Richard A. Posner, Observation, *The Economic Approach to Law*, 53 TEX. L. REV. 757, 765 (1975) (describing “the tendency of government to use queueing rather than pricing to ration access to the courts and other government services”).

387. See *supra* Part I.D.

sheets.³⁸⁸ This sort of accessibility hurdle again has distributional effects. Resources, whether fiscal or human, can be used to employ technologies—or people—to sort through poorly produced data and find the relevant sources. As we said in the context of data production and capture, Congress also should take an interest in the standardization of data at the access stage.

The question of how to choose definitively among these hurdles is beyond the scope of this Article. To be sure, policymakers should consider the constitutional values discussed in Part II in designing hurdles.³⁸⁹ But usually those values will not be dispositive. A more functional approach is required. Our analysis here suggests that policymakers should consider their choice of hurdles—and their choice between walls and hurdles—in light of the social value of different data uses. Sometimes, the best path will be to discriminate between high- and low-value users in a particular and specific fashion. Of course, this requires a democratic judgment as to what counts as an activity of social value. And this in turn needs to be determined either by an elected body (such as Congress) or an unelected one (such as an individual judge). Many decisions about social value are decentralized and out of the hands of elected actors. Under current policy, for example, PACER fee waivers are granted on a case-by-case, court-by-court basis.³⁹⁰ A debt-collection firm seeking to locate debtors is unlikely to receive a waiver. Scholars studying the effects of appointed counsel in prisoner litigation might get a waiver, although (unfortunately in our eyes) often they will not.³⁹¹

At other times, high-value users can be subsidized in bulk. Federal courts' disclosure of anonymized summary information is best understood in these terms, informing the public about how courts work and assisting with certain types of research.³⁹² Yet the existing range of such disclosures, in our view, does not go far enough. The inclusion of judge identifiers would promote social values—and it could be done costlessly by the courts, as judge identifiers are part of the CM/ECF and PACER records that are used in the IDB and elsewhere. A recent controversy around inter-judge sentencing disparities shows both the promise of disclosure and the perils of failing to do so systematically. In brief, an academic paper purported to show sentencing

388. *See id.*; *see also supra* note 166 and accompanying text.

389. *See supra* Part II.B.

390. Gelbach, *supra* note 46, at 339 (“Even though there is now a centralized process for requesting a fee waiver, the decision over whether to grant requested waivers is up to individual district courts” (footnote omitted)); *see also Fee Exemption Request for Researchers*, PACER, <https://perma.cc/FRW8-FD3Z> (archived Apr. 17, 2024) (allowing multi-court requests).

391. *See generally* Gelbach, *supra* note 46, at 338-41 (recounting examples of fee-waiver denials).

392. *See supra* Part I.C.2.

disparities among judges on the basis of defendant race, including identifying judges by name that had the largest disparities.³⁹³ On the one hand, even though this information is troubling, it seems to us vital for policymakers and the public to understand how racial disparities arise in the legal system. Only known disparities can be addressed. On the other hand, this paper had a major flaw—it relied on the incomplete and unreliable data from the courts.³⁹⁴ So there is some reason to worry that identified judges were erroneously flagged. The solution, we think, is *more* transparency—the federal courts should be concerned about *false* claims of racial disparities, and only improved data quality and access can bring the rue pattern of such disparities to light.

Critics of data disclosure also worry about the effect it might have on judges. For example, the Civil Justice Reform Act requires that, every six months, a list is published of every case that has been pending for more than six months and the name of the presiding judge.³⁹⁵ No financial or other compensation is tied to this list. Judges, with life tenure, retain their jobs no matter how long their list is. And yet, the so-called “Six-Month List” appears to have a substantial impact on how judges do their jobs.³⁹⁶ Case closures spike in the weeks preceding the publication of the list.³⁹⁷ This suggests judges can be incentivized to follow rules simply by publishing data about their behavior. But this experience also offers ground for caution. One study found “suggestive evidence that the List negatively affects accuracy.”³⁹⁸ The researchers explained, “The risk that the List may affect outcomes by increasing error arises from a possible tradeoff between speed and accuracy, a problem that was never addressed in the initial decision to implement this incentive.”³⁹⁹ “Teaching to the test,” so to speak, is as bad in the courthouse as in the classroom.

But even if disclosure of judge identifiers distorted judicial decisionmaking in the manner of the Six-Month List, we would be inclined to support that information subsidy as justified in terms of social welfare. Indeed, perhaps the effect of the Six-Month List is so strong because it is one of the few public

393. Charles Toutant, *Backpedaling: Authors of Study on Racist Rulings Retract Their Claims Against Pennsylvania, New Jersey Judges*, N.J. L.J. (Aug. 18, 2021, 5:18 PM), <https://perma.cc/TUY5-G5WS> (to locate, select “View the live page”).

394. *See id.* For the controversial study, with subsequent redactions, see Christian Michael Smith et al., *Racial Disparities in Criminal Sentencing Vary Considerably Across Federal Judges*, SOCARXIV, July 27, 2021, at 4-5, 17, <https://perma.cc/B6AC-Z9TS> (describing the database while acknowledging its incompleteness).

395. 28 U.S.C. § 476(a); *see, e.g., September 2023 Civil Justice Reform Act*, *supra* note 171.

396. *See de Figueiredo et al., supra* note 171, at 368-69, 390-92.

397. *See id.* at 392-97.

398. *Id.* at 371.

399. *Id.*

measures of its kind. To continue to our earlier analogy, if the test is sufficiently capacious, teaching to it is less of a problem.

C. The Possibilities of Proper Stewardship

We have suggested above a roadmap for the congressional management of judicial data as a public asset. This plan involves attention to data capture and production, with minimal limits and an emphasis on accuracy and uniformity. It also demands a presumption of data transparency, subject to exceptions (walls and hurdles) informed by constitutional values and public policy.

What we have not explained, or have not stated in full, are the plural goods that these efforts might produce. It would be impossible to catalog the innumerable ways that the public asset of judicial data might be used to serve the public. But we think it is important to highlight a few examples here for two reasons. First, we hope to show that the stakes here are not just academic (in the sense of being good for us academics). Rather, they involve public values and public rights. Second, we want to draw special attention to those public uses of judicial data currently being overlooked as support for our claim that legislation should do more than merely tweak the PACER fee structure. The dark data of federal courts can and should produce substantial public value. Our examples, then, underscore the idea that whatever data policy Congress and the courts adopt, the policymaking process should facilitate the realization of social utility of these underutilized resources.

1. Non-dispositions and access to justice

Access to justice is an important goal entangled in a thicket of difficult empirical questions. Counting the number of filings and the plaintiffs' win rate might provide some information about the role of federal courts in facilitating access to justice for different kinds of litigants.⁴⁰⁰ But much of the work necessary to ensure access to justice happens well before a case is decided.⁴⁰¹ Judicial data is a public good to the extent that it can contribute to our understanding of access-to-justice gaps and potential methods to redress them.

400. See, e.g., Lahav & Siegelman, *supra* note 327, at 1373-75 (finding an "approximately 40 percentage point (more than 50%) decrease" in plaintiff win rates in federal civil cases between 1985 and 1995, and raising both causal and normative questions about its significance).

401. See, e.g., Mitchell Levy, Comment, *Empirical Patterns of Pro Se Litigation in Federal District Courts*, 85 U. CHI. L. REV. 1819, 1822 (2018) (acknowledging that pro se reforms "may have improved the pro se litigation process by making it feel more humane and easier to understand and by giving litigants a stronger sense that their concerns have been heard," but arguing that the reforms have "been ineffective in improving case outcomes for pro se litigants, and alternative approaches should be considered").

For example, civil litigants who cannot afford to pay the filing fee may file motions to request in forma pauperis (IFP) status.⁴⁰² As Andrew Hammond has documented, there is no standard approach for determining IFP status among federal courts; each court may develop its own forms, and each judge make individual decisions about IFP applications.⁴⁰³ To extend Hammond's legal analysis, a team of researchers reviewed IFP applications for all cases filed in 2016, at a cost of over \$100,000 for PACER data.⁴⁰⁴ They determined that there was substantial inter-court and intra-court variation. For example, in one court, judges granted IFP applications at rates ranging from less than 20% to more than 80%.⁴⁰⁵

Similar studies could help identify trends in access to justice at stages of litigation before disposition. Dark judicial data would be invaluable in parallel studies about how criminal defendants fare when represented by private as opposed to publicly funded counsel and many other issues. And they would only be feasible if Congress mandated better and deeper access.

More directly, court data including non-dispositions could inform the work of lawyers and parties. Currently, those with access to commercial databases will get a fuller—and more accurate—picture of judicial behavior than those relying on public sources.⁴⁰⁶ A comprehensive, free, and usable database could help level the playing field between those parties with well-heeled counsel and those without.⁴⁰⁷ For example, Joanna Schwartz suggests that plaintiffs should pursue civil rights claims against municipalities for police departments' disregard of information included in lawsuits brought against them, but this only works if plaintiffs' lawyers have access to the complaints making those allegations.⁴⁰⁸ To be sure, masses of data can contribute to asymmetries because those with resources can make better use of the data than those without. But the current system, in which almost all useful information is "pay to play," can be improved. Free and accessible data can facilitate cooperation among legal services providers, such as public defenders' offices, that is impractical under the current regime. The increasing

402. Stephen M. Feldman, *Indigents in the Federal Courts: The In Forma Pauperis Statute—Equality and Frivolity*, 54 *FORDHAM L. REV.* 413, 413-14 (1985) (discussing the origins and terms of the statute allowing for such status).

403. Andrew Hammond, *Pleading Poverty in Federal Court*, 128 *YALE L.J.* 1478, 1497-1503 (2019).

404. Pah et al., *supra* note 21, at 135-36.

405. *Id.* at 135.

406. See Boyd et al., *supra* note 193, at 481-86 (documenting the effects of these differences).

407. See *id.* at 472-77 (showing differences between commercial services and fully public sources).

408. See generally Joanna C. Schwartz, *Monell's Quick Fix*, 125 *COLUM. L. REV.* (forthcoming 2025), <https://perma.cc/3L56-Y74E>.

availability of generative AI tools to examine these large datasets at relatively low prices only increases the value of such public-facing databases.

2. Non-dispositions and judicial performance

There is an entire industry of judicial analytics built around predicting how judges will rule on certain cases.⁴⁰⁹ These efforts, in other words, have been de facto privatized. But there are any number of noncommercial, public-interested reasons to care about other aspects of judicial performance. Judicial data is a public good to the extent it can contribute to these analyses and improve the delivery of justice.

Consider judicial caseloads. On occasion, Congress adds judgeships to the federal courts. One way to decide which districts should receive more judges is to determine which districts are the busiest. Indeed, when the federal courts make recommendations to Congress, they do so based on judicial workload.⁴¹⁰ But how should workload be measured? Counting case filings is one measure, but a crude one. It misses the highly variable judicial labor associated with different kinds of cases. The federal courts have a more nuanced approach, assigning “case weights” based on case categories using NOS codes.⁴¹¹ But this approach is also imperfect because NOS codes are a noisy signal given how they are imperfectly assigned in the first instance by parties. More fine-grained data is needed to determine which courts and which judges are busy based on an examination of judicial activity captured by the rates of different sorts of entries on docket sheets. Dark data that allows us to understand how much effort is exerted in different kinds of cases—for example by enabling the creation of measures of the extent and intensity of judicial activity on a per case basis, as distinct from litigant activity—may fill this gap.

409. See Engstrom & Gelbach, *supra* note 373, at 1031-35, 1059-71 (discussing “legal tech” and the legal profession, and taking a deeper look at predictive analytics applied to forum shopping); Patrick Flanagan & Michelle Hook Dewey, *Where Do We Go From Here? Transformation and Acceleration of Legal Analytics in Practice*, 35 GA. ST. U. L. REV. 1245, 1253-55 (2019) (describing some of the industry players).

410. For example, in March 2019, the Judicial Conference recommended creating sixty-five new district court judgeships and five new circuit court judgeships based on its assessment of judicial workload. BARRY J. McMILLION, CONG. RSCH. SERV., R45899, RECENT RECOMMENDATIONS BY THE JUDICIAL CONFERENCE FOR NEW U.S. CIRCUIT AND DISTRICT COURT JUDGESHIPS: OVERVIEW AND ANALYSIS 12, 15 (2019), <https://perma.cc/9HHJ-KLWH>; see also Elizabeth Wydra, Opinion, *There’s More to Repairing Federal Courts Than Supreme Court Expansion*, HILL (Feb. 10, 2021, 3:00 PM ET), <https://perma.cc/7STB-X88T>.

411. See *supra* note 355 and accompanying text; see also Philip Habel & Kevin Scott, *New Measures of Judges’ Caseload for the Federal District Courts, 1964-2012*, 2 J.L. & CTS. 153, 159-60 (2014) (discussing the case-weight system, including its 2004 revision and “events-based” approach).

Relatedly, the data that can be used to evaluate judicial performance also could be leveraged to evaluate attorney performance. The market for attorneys, especially attorneys for non-repeat litigants, is subject to large informational asymmetries.⁴¹² Potential litigants do not know which attorneys to hire. Once they hire an attorney, it is extremely difficult to monitor their performance.⁴¹³ Big data tools such as LLMs might permit such ex ante evaluation and ongoing monitoring, either by allowing for comparisons between attorneys or by offering predictions about how an equally skilled attorney would perform on a given set of facts.⁴¹⁴ But, again, such analysis depends on better availability of court data that Congress could mandate. Facilitating effective evaluation of attorneys may or may not be in the bar's interest—but it is plainly a public good latent in judicial data.

3. Court data and non-court alternatives

Courts are not the only means of delivering justice. Court alternatives could substitute in full—instead of going to court, parties would go to online dispute resolution instead.⁴¹⁵ Or they could serve as screening tools for potential litigants—first ask an artificial-intelligence (AI) tool the probability of winning a case or a motion and only invest in filing the papers or hiring a lawyer if the probability is sufficiently high. Judicial data may be a public good to the extent that it can be used to inform the creation of alternatives to court. One possibility is the use of AI to handle large volumes of cases.⁴¹⁶ A prerequisite for using machine learning, however, is the availability of a

412. See, e.g., Daniel L. Rubinfeld & Suzanne Scotchmer, *Contingent Fees for Attorneys: An Economic Analysis*, 24 RAND J. ECON. 343, 343 (1993) (identifying two types of information asymmetries between attorneys and clients). The landmark paper on selection problems in service markets is Kenneth J. Arrow, *Uncertainty and the Welfare Economics of Medical Care*, 53 AM. ECON. REV. 941 (1963).

413. See Larry E. Ribstein, *Ethical Rules, Agency Costs, and Law Firm Structure*, 84 VA. L. REV. 1707, 1707 (1998) (“Agency costs may arise when clients delegate discretion over their affairs to lawyers.”).

414. See David Restrepo Amariles, Pablo Marcello Baquero, Paul Boniol, Rajaa El Hamdani & Michalis Vazirgiannis, *Computational Indicators in the Legal Profession: Can Artificial Intelligence Measure Lawyers' Performance?*, 2021 U. ILL. J.L. TECH. & POL'Y 313, 315-17 (positing different computational indicators of attorney quality and the potential impacts of artificial intelligence).

415. See Shekhar Kumar, *Virtual Venues: Improving Online Dispute Resolution as an Alternative to Cost Intensive Litigation*, 27 J. MARSHALL J. COMPUT. & INFO. L. 81, 82-85 (2009) (documenting the use of such mechanisms and identifying four potential benefits); David Freeman Engstrom, *Digital Civil Procedure*, 169 U. PA. L. REV. 2243, 2273-83 (2021).

416. Cf. Aziz Z. Huq, *A Right to a Human Decision*, 106 VA. L. REV. 611, 634 (2020) (noting that “the state uses machine learning to make a range of high-volume decisions turning on empirical predictions for which a large pool of historical data is available”).

“training” set of data derived from previous interactions.⁴¹⁷ While federal court decisions are not free of errors or distortions,⁴¹⁸ they offer an obvious set of training materials for an AI system working as a substitute for federal courts. And because the work of federal courts is more than the issuance of final judgments, more court data should be input into these systems.

4. Courts and the social world

Federal court data also might be a public good in that it can help reveal facts about the social world that inform future public and private action. Imagine a policymaker interested in dealing with the problem of police or prison violence. They are obviously concerned with *where* such violence is concentrated and *what factors* predict its levels. It is possible (although not certain) that judicial data could be useful in identifying underlying patterns of constitutional violations by law enforcement or prison officials. Simply counting the number of lawsuits that name a police officer or prison employee as a defendant is a first step toward successful prediction.⁴¹⁹ But judicial data potentially allows deeper insights. For example, if data were accessible and easily analyzed, an algorithm might be able to review every complaint and identify not only when a suit is filed against a government official, but also the nature and frequency of the alleged wrongful conduct and the way in which the prisons’ lawyers respond. Insights might be gleaned from examining and comparing documentary discovery across cases. To be sure, not all complaints report the facts accurately, but they represent an untapped data source for policy reform.

Professional malpractice is another area in which this pattern analysis might be useful for regulators and consumers. One study found that doctors who have paid on a single medical malpractice claim have a fourfold higher risk of future claims than doctors who have never paid a claim.⁴²⁰ Although this finding can be interpreted in different ways, it might be understood to

417. See Aziz Z. Huq, *Constitutional Rights in the Machine-Learning State*, 105 CORNELL L. REV. 1875, 1886 (2020).

418. The training data used for all of these measures would embody historical biases on the part of federal judges, so an important question is whether and how the data could be corrected for such biases. For an example of a correction mechanism respecting racial bias, see Crystal S. Yang & Will Dobbie, *Equal Protection Under Algorithms: A New Statistical and Legal Framework*, 119 MICH. L. REV. 291, 293-97, 343-57 (2020).

419. In a similar vein, for evidence of how civilian allegations predict future civil rights litigation for police, see Kyle Rozema & Max Schanzenbach, *Good Cop, Bad Cop: Using Civilian Allegations to Predict Police Misconduct*, 11 AM. ECON. J. 225, 225-27 (2019) (finding the relationship to be “strong” and “nonlinear”).

420. Kowsar Yousefi, Bernard Black & David A. Hyman, *Paid Medical Malpractice Claims: How Strongly Does the Past Predict the Future?*, 20 J. EMPIRICAL L. STUD. 818, 837 (2023).

suggest that physicians who had made past payments in tort cases tend to be more likely to commit future tortious acts. Assuming that interpretation to be true, might the same phenomena hold for lawyers? Judicial data may be able to give further leverage on this question and related inquiries.

Importantly, court data provides information both where it is found and where it is not. The statistician David Hand has explored how identifying unexpected gaps in data can be informative for future decisionmaking.⁴²¹ A lack of cases where litigation might be expected could, for example, be evidence of an access-to-justice problem. At least in the context of public defense, scholars have attended to spatial inequities that arise when indigent defense is unevenly provided between different states and localities.⁴²² But there is much more to learn. Does the lack of environmental lawsuits in a particular district mean that there are fewer environmental problems or fewer environmental lawyers? Why are civil rights lawsuits filed at different rates across the country?⁴²³ Judicial data can help study these questions, though of course it is not sufficient on its own.

A related possibility is the use of judicial data to inform other government action, and thus to support the production of other public goods. For example, across a range of areas, federal law is substantially enforced through private litigation, sometimes called private enforcement.⁴²⁴ Securities, antitrust, consumer law, environmental law, and civil rights are areas significantly impacted by private enforcement.⁴²⁵ Some of the relevant statutes require private enforcers to give notice to the relevant government agency.⁴²⁶ But not all do.⁴²⁷ In the latter areas, federal agencies interested in knowing the landscape of private enforcement must pay for PACER or commercial data like the rest of us. Indeed, even in areas with notice, government lawyers must pay to review the relevant documents unless they are served by one of the parties. To the extent that private enforcement informs agency decisions about public

421. HAND, *supra* note 17, at 3-12.

422. See, e.g., Lisa R. Pruitt & Beth A. Colgan, *Justice Deserts: Spatial Inequality and Local Funding of Indigent Defense*, 52 ARIZ. L. REV. 219, 222-23 (2010) (noting that indigent defense “systems vary from state to state and even from county to county within a given state”).

423. For an analysis that considers both bench and bar influences, see Joanna C. Schwartz, *Civil Rights Ecosystems*, 118 MICH. L. REV. 1539, 1552-55, 1557-59 (2020).

424. Stephen B. Burbank, Sean Farhang & Herbert M. Kritzer, *Private Enforcement*, 17 LEWIS & CLARK L. REV. 637, 673 (2013).

425. *Id.*

426. E.g., 33 U.S.C. § 1365(b)(1) (Clean Water Act); 42 U.S.C. § 7604(b)(1) (Clean Air Act).

427. E.g., 35 U.S.C. § 292 (2010). *But see* 35 U.S.C. § 292 (2011) (amending the statute to disallow private enforcement).

enforcement or public regulation,⁴²⁸ federal court data may be important for furthering federal policies across a range of areas.

Similarly, many public entities are themselves regulated through litigation. The resulting lawsuits can generate important information for the management of these entities. For example, writing about civil rights litigation, Margo Schlanger and Denise Lieberman observe that “[c]ourt decrees have long been a source of regulatory innovation—and even organizations and jurisdictions that have not been sued often want to know what those decrees provide, to consider whether similar practices and procedures might be useful for them.”⁴²⁹ And even if solely for instrumental reasons, prisons or police forces should (and occasionally do) want to know more about litigation patterns that affect their work.⁴³⁰

Conclusion

Judicial data is valuable in governance and for those interested in understanding how they are being governed. In the absence of hurdles or walls, it can illuminate the operation of one branch of the national government. Constitutional lapses can be brought to light through retail or wholesale data. Such data, if made available as an information commons, can serve as a public good: It is a resource for researchers and reformers aiming at a better understanding, or an improvement in, judicial institutions. Or it can be privatized in whole or part to work as a source of profit either for the judiciary itself or its commercial adjuncts. This might be done without a wall that keeps all public eyes out; it might also be done by surrounding public databases with hurdles, in the form of “kludgy” interfaces, by limiting their amenability to search, or simply by requiring downloads to proceed page-by-page and at a fee.

These design choices, however, have remained largely out of sight. The informational economy of the judiciary is a “dark” zone for the simple reason it has not been isolated and studied. Yet it is also a field of distinctive institutional, democratic, and constitutional importance. Our first

428. On the relationship between public and private enforcement, see, for example, Zachary D. Clopton, *Redundant Public-Private Enforcement*, 69 VAND. L. REV. 285, 291-99 (2016) (documenting the pervasiveness of overlapping and redundant regimes); Aziz Z. Huq, *The Private Suppression of Constitutional Rights*, 101 TEX. L. REV. 1259, 1269-95 (2023) (documenting the historical pervasiveness of delegated, private enforcement mechanisms).

429. Margo Schlanger & Denise Lieberman, *Using Court Records for Research, Teaching, and Policymaking: The Civil Rights Litigation Clearinghouse*, 75 UMKC L. REV. 155, 161-62 (2006).

430. See Joanna C. Schwartz, *What Police Learn from Lawsuits*, 33 CARDOZO L. REV. 841, 891 (2012) (“Despite widespread reluctance to pay attention to litigation data, law enforcement agencies can—and do—learn from lawsuits.”).

contribution in this Article has been to start filling in what is happening in that dark zone, both as a matter of on-the-ground practice and of law. We have hence presented a positive, descriptive account of judicial data, emphasizing presently obscure corners of judicial data production and capture. We have also offered a much-needed doctrinal analysis of how judicial data flows and the likely constitutional limits on its regulation. This constitutional analysis set the stage further for a taxonomy of policy challenges and recommendations to expand public access to judicial data in furtherance of the public good.

We live in a moment in which information is increasingly the currency of the realm. The federal courts sit athwart a veritable gold mine of such information—and do not leverage it either for the general public good, or even to provide “the just, speedy, and inexpensive determination of every action and proceeding.”⁴³¹ The scholarly failure to consider how judicial data will be distributed; whether it will lead to a more effective and more equitable judiciary; and whether it will yield public goods or private profit is equally problematic. Acting upon this resource requires democratic action by Congress. At a moment when the national legislature seems capable of little more than pointless bickering, this may seem like asking a lot. But the failure of our third branch to deliver the quality of justice that the American people deserve will not be cured by exhortation alone. It needs data to get the job done.

431. FED. R. CIV. P. 1.