14

Free PACER

Jonah B. Gelbach

Data can fuel innovation. The next generation of legal scholars will stand not just on the shoulders of their forbears, but – hopefully – on the structural foundation created by access to the electronic public record. When they "surf to the courthouse door," they should find it open.[1]

This is the worst of the best of times for federal court data. Docket reports are compiled electronically, and the overwhelming number of documents are filed electronically using the CM/ECF system. Litigants, attorneys, and members of the public can obtain case information using searches of the Judiciary's PACER system, which stands for Public Access to Court Electronic Records. Not so long ago, none of these things was possible; people had to travel to the courthouse and conduct or request laborious searches for case information. But access is so much less than it could be, and this sorry state of affairs is the result of deliberate policy choices. We should free PACER's data, making federal litigation data free to the same public whose judicial system creates and stores those data.

Much of my case is based on a general argument about the value of providing free public access, especially to the scholars and journalists whose analysis can help monitor and improve our justice system. But there's a tech-specific case, too. Freeing PACER would broaden access to data, possibly allowing the development of accessible, affordable legal technology for the "have-nots" as well as those without the budgets necessary to collect or purchase their own data.[2]

Still, my focus in this chapter is on two other aspects of access to justice. First, the public's access to the way the justice system operates: access for those who measure,

[1] Proposed Brief Amicus Curiae for The American Association of Law Libraries, In Support of Plaintiffs' Motion for Summary Judgment at 13, Nat'l Veterans Legal Servs. Program, v. United States, 291 F. Supp. 3d 123 (D.D.C. 2018) (No. 1:16-CV-00745).

[2] For consideration of the distributional effects of legal tech in general, *see* David Freeman Engstrom & Jonah B. Gelbach, *Legal Tech, Civil Procedure, and the Future of Adversarialism*, 169 U. Pa. L. Rev. 1001 (2021).

https://doi.org/10.1017/9781009255301.019 Published online by Cambridge University Press

monitor, and evaluate the justice system as a whole. The second aspect is society's access to data-driven improvements in the justice system. Without access to large amounts of litigation data, society will forgo the gains from having scholars and journalists assess our judicial system's performance; anecdotes and surmise will rule the roost. I take as given that scholarship and journalism can help us evaluate and improve the justice we mete out. Thus, better data access gives us access to better justice by allowing us to better understand, better design, and better operate our justice system. For these reasons, this essay will focus on the case for freeing access to court data, especially for scholars.[3]

My argument is mercifully simple: The various policies and practices that have limited access to PACER for the past three decades have, as the Supreme Court once put it in an unrelated context, earned their retirement.[4] PACER began as an experimental dial-up system in 1988 before taking flight on the World Wide Web ten years later. It's been a long flight, actually. Though PACER has radically expanded public access compared to the status quo ante a third of a century ago, PACER searches cost too much for users and deliver too little. Its search functionality is roughly the same as it was back when. You can't search the text of docket entries, for example, much less the contents of filed documents. With some useful exceptions, PACER keeps our judicial system's public records behind a kludgy and unjustifiably high paywall.

PACER costs so much because in 1988, when Congress allowed the Judiciary to provide electronic public access, it directed the Judiciary to fund the initiative through user fees.[5] Happily, Congress could fix that problem with simple statutory text, and the fee revenues are so small relative to federal spending[6] that Congress needn't give a thought to whether it would break a sweat thinking about the question. Congress should mandate free PACER and cut the Judiciary a check large enough to replace the lost revenue. Problem solved.

* * *

[3] This essay focuses on federal court data and on scholars' access because those are the data and the users' needs I know best. That said, I believe the case for open access to court data is no weaker with respect to state and local courts, or with respect to journalists' access.

[4] *Cf.* Bell Atlantic Corp. v. Twombly, 550 U.S. 544, 563 (2007) (declaring that "after puzzling the profession for 50 years, [*Conley v. Gibson*'s] famous observation has earned its retirement").

[5] *Frequently Asked Questions*, Pub. Access to Ct. Elec. Recs., https://pacer.uscourts.gov/help/faqs/why-does-pacer-charge-fee. The Judiciary's current obligations and discretion in this sphere are creatures of the E-Government Act of 2002; for background, *see, e.g.*, Stephen J. Schultze, *The Price of Ignorance: The Constitutional Cost of Fees for Access to Electronic Public Court Records*, 106 Geo. L.J. 1197 (2018). Whether PACER's fees fit within the scope of that statute is a matter of ongoing controversy, with litigation pending about whether PACER fees are unlawfully, or just indefensibly, high; various lawsuits have been filed related to this question. One case was recently remanded to the district court after the Court of Appeals affirmed the district court's denial of cross motions for summary judgment. National Veterans Legal Servs. Program v. United States, 968 F.3d 1340 (Fed. Cir. 2020).

[6] *See* discussion in Section 14.3, *infra*.

https://doi.org/10.1017/9781009255301.019 Published online by Cambridge University Press

Section 14.1 of this chapter provides a brief discussion of PACER's history and present. In Section 14.2 I argue that opening PACER up to the public, with a focus on scholarly access, would lead to a greater quantity of research on the federal courts while also allowing scholars to carry out higher-quality studies.[7] It is time to unlock the doors to the benefits of open research on our justice system. Section 14.3 offers a vision for the future, centered on free access to PACER and direct funding of Judiciary activities to replace lost fee revenues, as well as fostering an attitude of openness in the Judiciary.

It's time to take down the paywall that divides judicial records and the public.

14.1 DIAGNOSING SCHOLARSHIP'S PACER PREDICAMENT

Litigation in most federal courts is managed through the Case Management/Electronic Case Files (CM/ECF) system operated by the Administrative Office of the US Courts (AO).[8] Parties file pleadings, motions, and other case documents through CM/ECF; the Court also uploads documents to this system, and textual information about case status, scheduling, and so on, exists as docket entry text. PACER is an online system that allows members of the public to get case information, including both docket reports that list case activity and case documents.[9]

[7] I have expressed some skepticism about quantitative empirical studies' ability to answer interesting questions related to litigation, due to the fact that litigant behavior is endogenous to legal rules; *see, e.g.*, Jonah B. Gelbach, *Can the Dark Arts of the Dismal Science Shed Light on the Empirical Reality of Civil Procedure?* 2 STAN. J. COMPLEX LITIG. 223 (2014), and also argued that with the right combination of data and analytical thinking we can learn about causal effects of legal rules; *see, e.g.*, Jonah B. Gelbach, *Material Facts in the Debate over* Twombly *and* Iqbal, 68 STAN. L. REV. 369 (2016). Moreover, there are areas where even solid descriptive work would advance our understanding of federal litigation. *See, e.g.*, Jonah B. Gelbach & Deborah R. Hensler, *What We Don't Know about Class Actions but Hope to Know Soon*, 87 FORDHAM L. REV. 65 (2018).

[8] The Administrative Office of the United States Courts provides links to the CM/ECF system for the United District Courts, Judicial Panel on Multidistrict Litigation, Courts of Appeals, Bankruptcy Courts, Court of Federal Claims, and International Trade at *Court CM/ECF Lookup*, PUB. ACCESS TO CT. ELEC. RECS., https://pacer.uscourts.gov/file-case/court-cmecf-lookup. The US Supreme Court maintains its own case management system. *See* SUP. CT. OF THE U.S., https://www.supremecourt.gov/filingandrules/faq_electronicfiling.aspx (providing details).

[9] The AO's webpage describing PACER is PUB. ACCESS TO CT. ELEC. RECS., https://www.uscourts.gov/court-records/find-case-pacer. Information not uploaded via CM/ECF is not available. For example, federal courts implemented CM/ECF at different times. *See, e.g.*, Gelbach, *Material Facts in the Debate over* Twombly *and* Iqbal, at 396 n.71. Other documents may be unavailable if filed under seal or for privacy reasons. *See* PUB. ACCESS TO CT. ELEC. RECS., https://pacer.uscourts.gov/help/faqs/are-full-social-security-numbers-displayed-court-documents?page=3).

14.1.1 *PACER Basics*

PACER[10] began as a dial-up-based experiment in 1988 by the Judicial Conference,[11] at a time when the public had to go to physical courthouses for case information.[12] Because PACER isn't free to operate, and Congress doesn't appropriate funds for it,[13] the Judicial Conference has long charged fees for PACER use. Since 1998, PACER has been accessible through World Wide Web connections; the per-page fee for downloaded information is now $0.10.[14]

Such fees make large case-info downloads costly. By policy choice, the Judicial Conference limits fees in some ways; today, charges are capped at $3.00 per document.[15] In addition, PACER waives fees for any quarter in which a user incurs under $30 in charges.[16] Parties and their attorneys receive a free copy of any electronically filed document "if receipt is required by law or directed by the filer."[17] Judicial opinions are supposedly free for anyone's download,[18] although that may not always be true in practice.[19]

Unfortunately, these policies do little for scholars who must download many documents for their work. The Judicial Conference does allow courts to exempt some persons from fees, including "individual researchers associated with educational institutions,"[20] and exemptions have facilitated important research, but much less than broader scholarly access would allow.

[10] Much/all of this discussion comes from National Veterans Legal Services Program v. U.S., 968 F.3d 1340 (Fed. Cir. 2020). For more on PACER and CM/ECF, see Schultze, *The Price of Ignorance*; J. Michael Greenwood & Gary Bockweg, *Insights to Building a Successful E-Filing Case Management Service: U.S. Federal Court Experience*, Int'l J. for Ct. Admin., June 2012, at 2; John Brinkema & J. Michael Greenwood, *E-Filing Case Management Services in the US Federal Courts: The Next Generation: A Case Study*, Int'l J. for Ct. Admin., July 2015, at 3.

[11] *Nat'l Veterans Legal Servs. Program* ("The Judicial Conference, in 1988, first authorized an 'experimental program of electronic access for the public to court information.'") (citing Jud. Conf. of the U.S., Rep. on the Proc. of the Jud. Conf. of the U.S. (1988), https://www.uscourts.gov/sites/default/files/reports_of_the_proceedings_1988-09_0.pdf).

[12] *Nat'l Veterans Legal Servs. Program*.

[13] *Id.*

[14] *Id.* (citing *Electronic Public Access Fee Schedule*, U.S. Cts. (Dec. 31, 2019), https://www.uscourts.gov/services-forms/fees/electronic-public-access-fee-schedule).

[15] *Electronic Public Access Fee Schedule*, U.S. Cts. (Dec. 31, 2019), https://www.Uscourts.Gov/Services-Forms/Fees/Electronic-Public-Access-Fee-Schedule.

[16] *Id.*

[17] *Id.*

[18] *Id.*

[19] *See* 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 (2018).

[20] *See Electronic Public Access Fee Schedule*, U.S. Cts. (Dec. 31, 2019), https://www.uscourts.gov/services-forms/fees/electronic-public-access-fee-schedule. Other persons are also eligible for exemptions, e.g., indigent persons and pro bono attorneys. *Id.*

14.1.2 *PACER Serves Up Case Records in a Clunky Way*

What PACER allows users to access remains far too limited. Critically, as the Free Law Project notes: "There is no document-level search."[21] And although PACER users can search for litigants' names or retrieve a case using a known docket number, it is impossible to search, say, for some version of "all cases that have at least one docket entry with the string 'FAILURE TO STATE A CLAIM' or 'QUALIFIED IMMUNITY.'" The Free Law Project observes: "This makes the system utterly unusable for general purpose research."[22] Anyone seeking to obtain data from many cases of a type – those with a Rule 12(b)(6) motion, or involving qualified immunity – must pay for enormous numbers of documents. This drains research budgets for no good reason.

14.1.3 *Financial Facts for PACER*

The Judiciary receives no funding from Congress for PACER, and the CM/ECF backbone is supported entirely by PACER fees. Each year these fees bring in around $150 million.[23] To put it mildly, that greatly exceeds the costs to the Judiciary of operating PACER, which one source recently put at $3 million.[24]

Who uses PACER? The Judiciary emphasizes PACER's enormous user base – more than 3 million user accounts, by 2020.[25] As of 2015, 75 percent of PACER *usage* was unpaid due to fee waivers (perhaps principally the $30/quarter free-usage allotment), even as 75 percent of PACER *revenues* came from fees paid by just 1 percent of account holders.[26] Although I am aware of no precise data, it's easy to guess that the biggest paying users include the big three electronic legal data vendors – Westlaw, LexisNexis, and BLAW – and bankruptcy data providers that service claimants' data needs. As of 2015, 63 percent of PACER usage was

[21] Brian Carver, *What Is the "PACER Problem"?* Free L. Project (Mar. 24, 2015), https://free.law/2015/03/20/what-is-the-pacer-problem/.

[22] *Id.*

[23] *See, e.g.*, Judiciary Info. Tech. Fund, U.S. Cts., https://www.uscourts.gov/sites/default/files/fy_2020_jitf_0.pdf (providing 2020 estimate of $147.7 million in "Estimated Receipts and Prior Year Recoveries" for the revenue on "EPA Program," which includes PACER, in Table 11.1 at page 11.2).

[24] Matt Ford, *The Courts Are Making a Killing on Public Records*, New Republic (Jan. 31, 2019), https://newrepublic.com/article/153003/courts-making-killing-public-records-pacer-fees.

[25] *See Members of PACER User Group Selected*, U.S. Cts. (Jan. 9, 2020), https://www.uscourts.gov/news/2020/01/09/members-pacer-user-group-selected.

[26] These figures come from my notes taken at the Increasing Access to Federal Court Data Workshop held at Penn Law in Fall 2015, memorialized in Jonah B. Gelbach, Final Report and Summary of Workshop on Increasing Access to Federal Court Data (on file with author) (also noting that as of 2015 there were roughly 2 million PACER user accounts). Other information about the grant, including the required Project Outcomes Report, may be viewed at https://tinyurl.com/yxcnzgzd.

undertaken by law firms, 12 percent by litigants themselves, 10 percent by commercial users, and only 3 percent by academics.[27] That academics use PACER so little is likely a testament to the fact that the large data pulls we need to do serious quantitative research would cost so much that most scholars either never start or rapidly abandon plans to study federal litigation using PACER data; *see* discussion in Sections 14.1.4 and 14.2.2 for more.

These facts paint a simple picture. PACER's paywall brings some money to the Judiciary, allowing Congress to fund it a bit less. The wall doesn't stop the biggest legal data vendors from downloading a lot of data, which they make available to subscribers who pay for platform access. The wall likely doesn't stop litigants in most individual cases, at least not with respect to their own cases. Nor does it likely do much to keep interested small-hobbyists out.[28] But the PACER paywall sure *does* screen out scholars.

14.1.4 *PACER Is Prohibitively Expensive for Scholars*

To study litigation, the judicial system, and how they interact with social phenomena, scholars need large amounts of data. Because of PACER's high per-page download fees, even projects involving relatively modest data can be very expensive.[29] To wit, a recent PACER fee waiver request estimates that to download only the docket sheets for just two years' worth of FOIA cases would cost nearly $4,000.[30] And consider a commentary recently published in *Science*,[31] which presents information on thousands of cases in which a party sought *in forma pauperis* status and shows considerable variation within districts in the likelihood such a waiver would be granted. The authors downloaded every docket report for all cases filed in 2016, at a cost they describe as "more than $100,000."[32] That won't become a common research strategy.

Although some research is conducted with second-best data sources (*see* section14.2.2, *infra*), I've had numerous conversations over the years with scholars who abandoned research ideas they thought promising because PACER's prohibitive costs made it too costly to push forward.

[27] *Id.*

[28] It is possible that free PACER would attract new hobbyists interested in bigger data corpora, though.

[29] Discretionary fee waivers might make a difference, if they were reliably and liberally granted, but it appears they aren't. *See* Section 14.2.2.

[30] *In re* Application for Multi-Court Exemption from the Jud. Conf.'s Elec. Pub. Access (EPA) Fees, No. 5:20-mc-00134 (D. Vt. Dec. 7, 2020) (estimating each docket report to cost an average of $2, with roughly 1,900 cases).

[31] Adam R. Pay et al., *How to Build a More Open Justice System*, 369 Science 134 (2020).

[32] *Id.*

14.2 THE CASE FOR ACCESS TO COURT DATA

As a general matter, PACER's paywall is unfortunate and out of place in a democracy, but it's especially irrational as applied to scholars. Under the status quo, scholars contribute very little in the way of fee revenues, so there is close to nothing to lose from granting us access. On the other hand, we'd contribute a lot of free research if we were allowed access. And privacy risks seem limited given that all documents in question already may be downloaded by anyone who knows to look for them. In sum, right now the Judiciary and the public are getting nothing for something.

14.2.1 *On Principle, Court Operations Should Be Presumptively Open to the Public*

Publicity is the very soul of justice. It is to publicity, more than to everything else put together, that the English system of procedure owes its being the least bad system as yet extant, instead of being the worst.

– Jeremy Bentham[33]

It's a basic principle of a democratic society that government's operations should be open to public scrutiny. Otherwise, the citizenry cannot tell whether government is operating justly, speedily, or inexpensively.[34] Even appropriate scrutiny of one governmental branch by another can be limited without sufficient openness.

Constitutional and federal common law bases for public rights to observe judicial proceedings are most developed in the context of criminal trials,[35] but these are of course only one spot in the landscape of judicial proceedings. Of the nearly 80,000 federal criminal cases filed in 2018, only about 2 percent went to trial,[36] a number that isn't much different for civil cases. With so few trials, pretrial proceedings delineate the fairness and future incentive effects of federal litigation. Accordingly, whatever interests of publicness inhere in open trial proceedings also extend to pretrial ones, too.

[33] Garth Nettheim, *The Principle of Open Justice*, 8 U. TASMANIA L. REV. 25 (1984).

[34] *Cf.* FED. R. CIV. P. 1 (the Federal Rules of Civil Procedure "should be construed, administered, and employed by the court and the parties to secure the just, speedy, and inexpensive determination of every action and proceeding.").

[35] Peter W. Martin, *Online Access to Court Records – from Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 855, 857 (2008) (citing Richmond Newspapers, 448 U.S. 555, 570 (1980)).

[36] *See* John Gramlich, *Only 2% of Federal Criminal Defendants Go to Trial, and Most Who Do Are Found Guilty*, PEW RSCH. (June 11, 2019), https://www.pewresearch.org/fact-tank/2019/06/11/only-2-of-federal-criminal-defendants-go-to-trial-and-most-who-do-are-found-guilty/.

Professor Peter W. Martin has described five purposes for public access to litigation-related documents: (1) assuring fairness – both actual and perceived – of proceedings, (2) ensuring that the public's ability to assess and criticize judicial performance can be done on an informed basis, (3) educating the public about how litigation works, and – hopefully – inspiring confidence as a result, (4) enabling the "check[ing]" of the judicial system by the public, and (5) "[p]roviding an outlet for community 'concern, hostility, and emotion' in cases that are in the public eye".[37] At least the first four of these are implicated in the Judiciary's refusal to allow affordable public access to PACER for scholarly or journalistic research.[38]

In sum, as House Judiciary Committee Chairman Jerrold Nadler recently stated: "It is a disservice that in today's digital age, the public's access to public records in public proceedings is so resource-intensive and burdensome …. It is indefensible that the public must pay fees – and unjustifiably high fees, at that – to know what is happening in their own courts."[39]

14.2.2 *Important Research Could Be Done with Freer Access, and for Free to the Judicial System*

This section provides a brief, ad hoc review of quantitative empirical work that has been done using federal court data.[40]

The Limits of Non-PACER Data Sources. Even before PACER, there was much quantitative empirical research published using federal court litigation data. A chief alternate data source is the Federal Judicial Center's Integrated Database (IDB), "by far the most prominent" of data sources used for statistical research by legal scholars.[41]

[37] Martin, *Online Access to Court Records*, at 856–57 & n.5.

[38] For an excellent discussion, see Schultze, *The Price of Ignorance*.

[39] Chairman Nadler Statement for the Markup of H.R. 8235, the Open Courts Act of 2020, House Comm. on the Judiciary (Sept. 15, 2020), https://judiciary.house.gov/news/documentsingle.aspx?DocumentID=3346.

[40] To be sure, quantitative research varies in quality, and assessments can and should be done by those focused on relevant substantive questions. An example of a problematic study is Christian Michael Smith et al., *The Most Discriminatory Federal Judges Give Black and Hispanic Defendants At Least Double the Sentences of White Defendants* (July 27, 2021), https://osf.io/download/61006a1f0c4cba01ccbc7224/?version=1&displayName=Judge%20Discrim%20Manuscript-2021-07-27T20%3A18%3A39.619Z.pdf. After I and others pointed out methodological problems, the authors withdrew the paper's references to individual judges and retitled it. *See* Christian Michael Smith et al., *Racial Disparities in Criminal Sentencing Vary Considerably across Federal Judges*, SocArXiv (2021), https://osf.io/preprints/socarxiv/j2gbn/. Nevertheless, I take the view that added capacity to do research with administrative records from the civil justice system is a net plus.

[41] Frank B. Cross, *Comparative Judicial Databases*, 83 Judicature 248, 248 (2000).

https://doi.org/10.1017/9781009255301.019 Published online by Cambridge University Press

Scholarly work using the IDB has spanned the gamut of topic areas for years.[42] But the IDB has its flaws.[43]

So, when FJC researchers set out to study summary judgment practice at the request of the Advisory Committee on Civil Rules in the wake of the Supreme Court's 1986 summary judgment trilogy, these researchers analyzed docket sheets randomly selected in three district courts.[44] Subsequent FJC studies of summary judgment practice also used docket information.[45] This is particularly notable because summary judgment dispositions are the sort of issue of interest not just to scholars or elected officials, but also to the Judiciary. In fact, the sort of data the FJC, and Professor Burbank,[46] has used to study summary judgment is the primary source of information for judges doing rule-making. Yet as intensive as this work was, it typically involved studying few districts, or even just one at a time.

The CM/ECF system now contains all data necessary to enable nationwide studies, and by 2008 (if not sooner), the FJC was doing so in response to Advisory

[42] *See, e.g.*, Miguel F. P. de Figueiredo, Alexandra D. Lahav & Peter Siegelman, *The Six-Month List and the Unintended Consequences of Judicial Accountability*, 105 Cornell L. Rev. 363 (2020); William H. J. Hubbard, *Testing for Change in Procedural Standards, with Application to* Bell Atlantic v. Twombly, 42 J. Legal Stud. 35, 49–53, 60 (2013); Jason Scott Johnston & Joel Waldfogel, *Does Repeat Play Elicit Cooperation? Evidence from Federal Civil Litigation*, 31 J. Legal Stud. 39 (2002); Warren E. Agin, *Predicting Chapter 11 Bankruptcy Case Outcomes Using the Federal Judicial Center IDB and Ensemble Artificial Intelligence*, 35 Ga. St. U.L. Rev. 1093 (2019); Erica J. Hashimoto, *Defending the Right to Self-Representation: An Empirical Look at the Pro Se Felony Defendant*, 85 N.C. L. Rev. 423 (2007); Theodore Eisenberg & Margo Schlanger, *The Reliability of the Administrative Office of the U.S. Courts Database: An Initial Empirical Analysis*, 78 Notre Dame L. Rev. 1455 (2003).

[43] For one, it contains only data of a certain snapshot type – what can be reported at case beginning and termination. It can't, for example, tell us anything about summary judgment motions that don't terminate an action. Stephen B. Burbank, *Vanishing Trials and Summary Judgment in Federal Civil Cases: Drifting toward Bethlehem or Gomorrah?* 1 J. Empirical L. Stud. 591, 606 (2004). And Professor Gillian Hadfield's influential study provides compelling evidence to doubt the accuracy of even the disposition coding that the IDB does contain. 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 L. Stud. 705 (2004).

[44] Joe S. Cecil & C. R. Douglas, Fed. Jud. Ctr., Summary Judgment Practice in Three District Courts 1, 3 (1987).

[45] *See, e.g.*, Joe S. Cecil, Rebecca N. Eyre, Dean Miletich & David Rindskopf, Fed. Jud. Ctr., Trends in Summary Judgment Practice, 1975–2000 (2007), https://www.uscourts.gov/sites/default/files/summary_judgment_1975-2000.pdf; Joe S. Cecil, Dean P. Miletich & George Cort, Fed. Jud. Ctr., Trends in Summary Judgment Practice: A Preliminary Analysis (2001), https://www.law.umich.edu/facultyhome/margoschlanger/Documents/Resources/Sources_Datasets_American_Trial_Trends/Trends_Summary_Judgment_Practice_Preliminary_Analysis.pdf); Joe S. Cecil, Rebecca N. Eyre, Dean Miletich & David Rindskopf, *A Quarter-Century of Summary Judgment Practice in Six Federal District Courts*, 4 J. Empirical L. Stud. 861 (2007).

[46] *See* Burbank, *Vanishing Trials and Summary Judgment*.

Committee requests.[47] Due to PACER's paywall, though, outside researchers cannot generally do likewise. That means the Judiciary, and society more widely, are partly subject to the limits of FJC's resources, however impressive its ingenuity, in addressing matters of interest to rule-making.

Another area of great interest in which the IDB is limited is the pleading standard, whose alteration in *Bell Atlantic Corp. v. Twombly*[48] and *Ashcroft v. Iqbal*[49] was arguably the top focus of procedure scholars for a period of several years. As Professor Burbank wrote even before those cases, the IDB's limitations make it difficult, at best, to determine anything specific about Rule 12(b)(6) and 12(c) motions decided under the Rule 8(a)(2) pleading standard, including even "the number of such motions filed or decided."[50]

There are still other data sources for scholars studying the federal legal system, all with key limitations. For example, the US Sentencing Commission posts defendant-level data on sentences in federal criminal cases.[51] Interesting research has been conducted using these data, for example, on racial disparities and the federal sentencing guidelines.[52] But they obviously concern only criminal sentences. Judicial opinion databases are another source of data; in some cases these can be useful,[53] though it is well known that for many others this approach often brings bias given that only some opinions are published.[54] Another alternative is the Court Listener RECAP archive,[55] hosted by the Free Law Project, which contains

[47] *See* Memorandum from Joe Cecil and George Cort to Judge Michael Baylson (Aug. 13, 2008), https://www.fjc.gov/sites/default/files/2012/SuJuLRS2.pdf.

[48] 550 U.S. 544 (2007).

[49] 556 U.S. 662 (2009).

[50] Burbank, *Vanishing Trials and Summary Judgment*. That said, there are some authors that have used the IDB in an attempt to learn about the effects of *Twombly* and *Iqbal*. *See* Hubbard, *Testing for Change in Procedural Standards* (using the IDB only); William H. J. Hubbard, *The Effects of* Twombly *and* Iqbal, 14 J. Empirical Legal Stud. 474 (2017) (using the IDB together with data coded based on PACER downloads); Kevin M. Clermont & Theodore Eisenberg, *Plaintiphobia in the Supreme Court*, 100 Cornell Law Rev. 193 (2014).

[51] *See Individual Offender Datafiles, U.S. Sent'g Comm'n,* https://www.ussc.gov/research/data files/commission-datafiles#individual.

[52] *See, e.g.*, Joshua B. Fischman & Max M. Schanzenbach, *Racial Disparities under the Federal Sentencing Guidelines: The Role of Judicial Discretion and Mandatory Minimums*, 9 J. Empirical Leg. Stud. 729 (2012); Sonja B. Starr & M. Amrit Rehavi, *Racial Disparity in Federal Criminal Sentences*, 122 J. Pol. Econ. 1320 (2014).

[53] *E.g.*, Theodore Eisenberg & Geoffrey P. Miller, *Attorney Fees in Class Action Settlements: An Empirical Study*, 1 J. Empirical Leg. Stud. 27, 44 (2004) (using Westlaw and Lexis searches to generate a data set of cases with class action settlements); *see also* Eric Helland & Jonathan Klick, *The Effect of Judicial Expedience on Attorney Fees in Class Actions*, 36 J. Legal Stud. 171 (2007) (using the Eisenberg & Miller data as one source in studying whether judges with crowded dockets are more likely to approve settlements).

[54] Burbank, *Vanishing Trials and Summary Judgment*.

[55] *Advanced RECAP Search, Ct. Listener*, https://www.courtlistener.com/recap/. For a colorful history of the development of RECAP, see Stephen Schultze, *PACER, RECAP, and the Movement to Free American Case Law*, VoxPopuLII (Feb. 3, 2011), https://blog.law.cornell .edu/voxpop/2011/02/03/pacer-recap-and-the-movement-to-free-american-case-law.

https://doi.org/10.1017/9781009255301.019 Published online by Cambridge University Press

"millions of PACER documents and dockets"[56] and grows over time as more users use RECAP, even as users are able to avoid paying for documents that RECAP already has.[57] As great a resource as RECAP is for documents that have been stored, there is no general way to measure how representative RECAP's sample of documents and docket reports is.[58]

PACER's Unrealized Potential. This brings me to work using all of PACER, or representative samples. Bankruptcy has been a prolific substantive area for empirical study, due in part to the enterprising efforts of Professors Teresa Sullivan, Elizabeth Warren, and Jay Westbrook. The challenges surmounted in this work are described engagingly by Professor Lynn LoPucki. In brief, the professors flew themselves and some photocopiers around the country, copied court files in the clerks' offices, and sent the copies to Austin, Texas, where they coded the information they would use – even though copy machines in the clerks' offices were unused after hours.[59] In 2001, after PACER's advent, the same researchers were able to secure judicial permission to hire "moonlighting assistant clerks to photocopy the files and ship the copies," because that allowed them to avoid PACER's per-page download fees.[60] As Professor LoPucki writes, "With the political constraints removed by judges, the 1981 technology – photocopying in the clerks' offices – was cheaper than the 2001 technology – internet PACER – with its one-page-at-a-time political restriction."[61] In sum, at enormous and duplicative cost, these researchers were able to obtain substantial data on bankruptcy cases. They used these data to publish numerous studies about an issue of great public policy concern.[62] Subsequent research on consumer bankruptcy reform was able to use both PACER fee waivers[63] – obviously a welcome development – and help from AACER, the Automated Access to Court Electronic Records, a service that collects bankruptcy court data on a daily basis and repackages

[56] *Advanced RECAP Search*, *Ct. Listener*, https://www.courtlistener.com/recap/ (public landing page encouraging visitors to "Search our database of millions of PACER documents and dockets").

[57] RECAP has some additional, though limited, docket information provided by federal courts via RSS feeds. Rebecca Fordon, *RSS Feeds, PACER, and the Fight for Access to Federal Docket Information*, RIPS L. Libr. Blog (Sept. 16, 2020), https://ripslawlibrarian.wordpress.com/2020/09/16/rss-feeds-pacer-and-the-fight-for-access-to-federal-docket-information/.

[58] That has not stopped thoughtful researchers from using RECAP data for projects in which representativeness may not be critical. *See, e.g.*, Christina L. Boyd & David A. Hoffman, *The Use and Reliability of Federal Nature of Suit Codes*, 2017 Mich. St. L. Rev. 997.

[59] Lynn M. LoPucki, *The Politics of Research Access to Federal Court Data*, 80 Tex. L. Rev. 2161, 2166–67 (2001).

[60] *Id.*

[61] *Id.*

[62] These include both books and scholarly articles. *See, e.g.*, Teresa A. Sullivan, Elizabeth Warren & Jay L. Westbrook, As We Forgive Our Debtors (1989); Elizabeth Warren & Jay L. Westbrook, *Financial Characteristics of Businesses in Bankruptcy*, 73 Am. Bankr. L.J. 499 (1999).

[63] Robert M. Lawless et al., *Did Bankruptcy Reform Fail? An Empirical Study of Consumer Debtors*, 82 Am. Bankr. L.J. 349, 354 (2008).

it commercially.[64] AACER must itself pay the Judiciary to download data that are already in the PACER system, but the facts of the bankruptcy system evidently make it possible for market forces to sustain AACER's more rational approach to data distribution. There is little prospect for such a market solution to PACER's paywall for nonbankruptcy cases (and would-be users must still pay AACER).

Other scholarly work has been published using the fruits of PACER fee waivers. Examples include work by Professor David Engstrom,[65] Professor Gillian Hadfield,[66] and Professor Alexander Reinert.[67] But PACER fee waivers are at most a partial solution to the PACER paywall, and a highly imperfect one at that. Even though there is now a centralized process for requesting a fee waiver,[68] the decision over whether to grant requested waivers is up to individual district courts.[69] This can extend the time and uncertainty related to project feasibility. It results from the fact that the CM/ECF system underlying PACER is really an amalgamation of individual systems, each run by an individual court as it sees fit. It is hard to think of good reasons why one court would deny fee waivers for the same scholarly research another approves, yet that happens.[70] Further, my conversations with other scholars indicate the possibility of inconsistent treatment. For example, although Professor Lawless and his coauthors received fee waivers from all but the Southern District of Texas,[71] one scholar at a workshop on federal court data access that I organized in 2015[72] reported being unable to receive fee waivers from more than thirty-five of the ninety-four district courts.[73] Study of the Supreme Court's alteration of the pleading

[64] *Id.*

[65] *See* David F. Engstrom, *Public Regulation of Private Enforcement: Empirical Analysis of DOJ Oversight of* Qui Tam *Litigation under the False Claims Act*, 107 Nw. U. L. Rev. 1689 (2013). Professor Engstrom confirmed to me that he obtained such waivers for this paper and two companions.

[66] Hadfield, *Where Have All the Trials Gone?*

[67] Alexander A. Reinert, *Measuring the Impact of Plausibility Pleading*, 101 Va. L. Rev. 2117 (2015).

[68] *Application for Multi-Court Exemption from the Judicial Conference's Electronic Public Access (EPA) Fees*, Jud. Conf. of U.S., https://pacer.uscourts.gov/sites/default/files/files/MultiCourt%20Exemption%20Request.pdf. Previously, exemptions had to be requested court by court. The multi-court exemption application mechanism is a recent development, perhaps the result of a constructive response by the AO to feedback from scholars. *See* Project Outcomes Report.

[69] *See, e.g.*, Order Granting PACER Waiver, *In re* Application for Multi-Court Exemption from the Jud. Conf.'s Elec. Pub. Access (EPA) Fees, No. 5:20-mc-00134-gwc (D.Vt. Dec. 7, 2020), ECF No. 2.

[70] David Freeman Engstrom, *The* Twiqbal *Puzzle and Empirical Study of Civil Procedure*, 65 Stan. L. Rev. 1203, 1248 (2013) (recounting one judge's statement that the relevant district court had a "long-standing policy not to grant exemptions to [PACER fees] for research" and characterizing such a "categorical unwillingness . . . to support empirical research designed to improve the administration of civil justice [a]s an embarrassment that calls out for correction by Congress or the Judicial Conference").

[71] Lawless et al., *Did Bankruptcy Reform Fail?* at 354 n.23.

[72] Increasing Access to Federal Court Data Workshop.

[73] Gelbach, *Final Report and Summary of Workshop*.

standard in *Twombly* and *Iqbal* also shows PACER's potential, and the limitations of its alternatives.[74]

Another anecdote involves a July 2020 request by Plymouth State University Professors Chantalle Forgues and Daniel Lee for PACER fee waivers so they could download docket reports, complaints, and petitions for all New England district courts.[75] The scholars planned to use computer programs to target a subset of cases as classified using Nature of Suit codes.[76] All district courts petitioned granted the request except for the District of Massachusetts.[77] Events over several months culminated in a court order denying the professors' request,[78] because of the Chief Judge's belief that the researchers planned to use software that, "according to our IT security personnel, would present an unacceptable degree of risk to our system."[79] Professors Forgues and Lee explained that "the script that petitioners propose is used everyday, including on weekends, to download files from PACER. The script at issue is used by big law firms and larger universities on a regular basis, and has been used on the Court's system consistently for over seven (7) years."[80] A motion seeking reconsideration remains pending.[81]

The upshot of all of this is two-fold. First, bulk downloads are available, without limit, to users who can afford to pay PACER fees, even as federal courts create situations in which researchers who cannot afford PACER fees are out of luck. More generally, the PACER system, and judicial policies and practices around it, stymie what would otherwise be PACER's staggering research potential. PACER contains

[74] *See* text accompanying note 50. See also reports including Joe S. Cecil et al., Fed. Jud. Ctr., Motions to Dismiss for Failure to State a Claim after *Iqbal* (2011), https://www.uscourts.gov/sites/default/files/motioniqbal_1.pdf; Joe S. Cecil et al., Fed. Jud. Ctr., Update on Resolution of Rule 12(B)(6) Motions Granted With Leave to Amend (2011), http://masonlec.org/site/rte_uploads/files/Gelbach_FJC%2012b6.pdf, which were based on access to PACER. These reports provoked considerable debate given the controversy related to the Supreme Court's actions in *Twombly* and *Iqbal*. Unfortunately, despite indications that it would release the underlying data, to my knowledge the FJC has never released even the docket numbers of the cases studied. That hamstrung other researchers and meant that further work would be delayed and costly in time and/or money. *See, e.g.*, Reinert, *Measuring the Impact of Plausibility Pleading* (performing an independent analysis of the issues and courts studied by the FJC, using PACER fee waivers to allow him to obtain data, but ultimately being able to obtain fee waivers for only fifteen of the twenty-three district courts the FJC included).

[75] Petitioners' Memorandum in Support of Motion to Alter or Amend Judgment, or in the Alternative Motion for Relief from Judgment or Order at 1, *In re* Application for Exemption from Elec. Pub. Access Fees for Chantalle R. Forgues, Daniel Lee, and Plymouth State University, No. 20-mc-91510-FDS (D. Mass. Dec. 31, 2020), ECF No. 7.

[76] *Id.*

[77] *Id.* at 1–2.

[78] *Id.*

[79] *Id.* at 6 (quoting Letter from F. Dennis Saylor IV, Chief J. U.S. Dist. Ct. D. Mass. (Dec. 4, 2020)).

[80] *Id.* at 7.

[81] *Id.* at 10.

https://doi.org/10.1017/9781009255301.019 Published online by Cambridge University Press

over a billion documents,[82] and that number grows daily. Many other federal agencies have made their data available, with privacy restrictions, allowing the public to benefit from enormous amounts of professional-quality research at no cost to the federal government.[83] The many examples of research noted above – from enterprising bankruptcy scholarship to illuminating work on summary judgment and pleading to work on trials – hint at the possibilities presented by piercing PACER's paywall. This research, much of it using inferior or limited sources of data, is a drop in the bucket compared to what we could have with free PACER.

14.2.3 *What about Privacy?*

A natural basis for concern about free PACER involves privacy of both litigants and witnesses. This essay is not the place to fully develop either the bases for such concerns, which are real, or the responses to them. But engaging these topics can allay certain aspects of these concerns while informing further inquiry as to others.

A first concern involves risks in the criminal justice system. Stephen Schultze chronicles one example:

> In 2006, the Department of Justice (DOJ) observed a disturbing new phenomenon: anonymous individuals had begun to use PACER to cull and republish witness information. DOJ stated in a letter to the Judicial Conference, "we are witnessing the rise of a new cottage industry engaged in republishing court filings about cooperators on websites such as www.whosarat.com for the clear purpose of witness intimidation, retaliation, and harassment." DOJ urged the Judiciary to suppress public access to all plea agreements.[84]

What DOJ wanted to preserve is what is sometimes called "practical obscurity" – "print-era barriers to access to sensitive criminal information."[85] After describing how federal judges engaged these questions, including at the level of the Judicial Conference's Court Administration and Case Management Committee, Schultze observes that "by the time whosarat.com appeared in newspaper headlines, the Judicial Conference had codified its approach to sensitive information in electronic court records," with the result that the Justice Department "could protect sensitive plea agreements by sealing or redacting the documents, but it could not rely on the Judiciary to recreate anachronistic approximations of practical obscurity."[86]

If this balance between privacy concerns and public access is good enough for those who might be at risk of physical violence, it is difficult to see why it should not

[82] *See* Sup. Ct. of U.S., 2014 Year-End Report on the Federal Judiciary 6 (2014), https://www.supremecourt.gov/publicinfo/year-end/2014year-endreport.pdf.
[83] Some of that research may be supported by public grant funding, but plenty of it is not.
[84] Schultze, *The Price of Ignorance*, at 1206.
[85] *Id.*
[86] *Id.*

also be good enough for others in the justice system. This is especially true in the civil context, where concerns related to personal embarrassment or trade secret interests can be handled on a case-by-case basis. Just as the Freedom of Information Act contains exceptions for trade secrets and certain other delicate information,[87] so, too, may federal courts allow litigants to file documents under seal in appropriate circumstances.

The privacy case for limiting access to PACER records en masse seems especially weak given that the very same records are *already available electronically*. To the extent that practical obscurity exists now, it is thus limited only by PACER fees. The paywall does not protect anyone from PACER's large commercial buyers, including, for example, bankruptcy-related services such as AACER. But these fees are high enough to keep out scholars, most of whom have little to no interest in individual information – the point of much scholarly research being to generalize rather than investigate details of unknown persons. It also includes journalists, which places PACER in tension with the First Amendment values that animate the constitutional basis for open court records that the Supreme Court found in *Richmond Newspapers, Inc. v. Virginia*.[88]

14.3 A VISION FOR THE FUTURE

I'm hardly the first one to advocate dismantling PACER's paywall. More than a decade ago, Professor LoPucki argued that eliminating the paywall "would open a real-time window on court-system operation superior to any previously possible," with the happy result that "[n]ot only researchers, but parties, lawyers, the government, and the public could see every important aspect of how the courts operate."[89] The Free Law Project has long argued that there is a "PACER problem."[90] Other scholars are on the case, in both academic journals and popular media.[91] Litigation that could reduce the magnitude of PACER fees – shorten the paywall – has had

[87] 5 U.S.C. § 552(b)(4); *see also* Christopher S. Yoo, Protected Materials in Public Rulemaking Dockets 113 (2020), https://www.acus.gov/sites/default/files/documents/Christopher%20Yoo%20FINAL%20REPORT%2011%2023%202020.pdf (discussing balancing of public disclosure interests against privacy interests).

[88] Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555 (1980). For more on *Richmond Newspapers*, and on privacy issues related to the case for freeing PACER, *see* Schultze, *The Price of Ignorance*.

[89] LoPucki, *Court-System Transparency*, at 537.

[90] Carver, *What Is the "PACER Problem"?*.

[91] *See, e.g., Hearing before the U.S. House of Representatives Judiciary Committee Subcommittee on Courts, Intellectual Property, and the Internet*, 116th Cong. (2019) (statement of Michelle Cosby, President of the American Association of Law Libraries), https://www.aallnet.org/wp-content/uploads/2019/09/AALL-Statement-for-HJC-Hearing-Record-092619.pdf; Sarath Sanga & David Schwartz, *Tear Down This Judicial Paywall*, Wall Street J. (Dec. 13, 2020), https://www.wsj.com/articles/tear-down-this-judicial-paywall-11607900423.

some success and is ongoing.[92] And as I discuss momentarily, there is a real chance that Congress will finally act to free PACER.

But what should a freed PACER regime look like? My purpose in this Section is to provide some suggestions. If anything like the Open Courts Act makes it into law, then some of these suggestions will have been either taken or eclipsed. Others would still be up for consideration if, as seems likely, Congress left the AO with substantial discretion in choosing how to make court records publicly available. Some of the possibilities I raise are concrete. The last section, though, is a meditation on the need for the Judiciary to adopt a new institutional attitude favoring openness. That could go a long way.

14.3.1 *Free PACER and Replace Its Fees with General Revenue*

Encouragingly, in December 2021 the Open Courts Act of 2021 was reported out of the Senate Judiciary Committee with bipartisan support.[93] This bill would require all court documents to be presumptively freely accessible to the public within three years after enactment.[94] The bill also calls for modernization of PACER's feature set,[95] and a similar bill is pending in the US House as of this writing.[96] Although the costs of modernization might be offset partly by congressional appropriations,[97] future funding would need to come from fees paid by federal agencies for their PACER use, based on inflation-adjusted fees agencies paid in 2018,[98] and possibly also increases in filing fees on parties.[99]

I have doubts about this approach to funding. First, I have not been able to find a breakdown of PACER fees by source, but logic indicates that fees paid by federal agencies are only part of the revenue from PACER. Thus, without increasing filing fees, the Judiciary would absorb at least some revenue loss. My goal is to free PACER, not cut courts. Second, if the Judiciary exercised the authority the Open Courts Act provides to raise filing fees, that burden would fall on litigants. My goal is to free PACER, not raise additional financial barriers to courts.

The better course is for Congress to simply (1) eliminate the Judiciary's discretion to charge PACER fees and (2) increase general revenue funds for the Judiciary.

[92] Nat'l Veterans Legal Servs. Program v. United States, 968 F.3d 1340 (Fed. Cir. 2020).
[93] *See* Nate Raymond, *Free PACER? Bill to End Fees for Online Court Records Advances in Senate* (December 9, 2021), https://www.reuters.com/legal/legalindustry/free-pacer-bill-end-fees-online-court-records-advances-senate-2021-12-09/.
[94] Open Courts Act of 2021, S. 2614, 117th Cong. § 3(a) (2021).
[95] *Id.* § 2.
[96] Open Courts Act of 2021, H.R. 5844, 117th Cong. (2021).
[97] Raymond, *Free PACER?*
[98] S. 2614, § 3(d)(1).
[99] *Id.*

https://doi.org/10.1017/9781009255301.019 Published online by Cambridge University Press

Congress could hold the Judiciary harmless, replacing all PACER paywall revenues, for just $150 million.[100]

Although $150 million would be a lot of money to have for yourself, it's a teeny-tiny little droplet in congressional terms. By way of comparison, Americans spend roughly twenty-five times as much on wild bird food as PACER fees brought in.[101] And overall congressional funding to the Judiciary in 2020 was fifty times what the PACER paywall brought in.[102]

A situation in which we have hamstrung public inquiry into the functioning of the federal litigation system over a literal rounding error may not be funny, but it really is silly. Let's stop it already. Congress should just cut an annual check to the Judiciary, allowing the public to better understand – and better inform Congress – how one of the three branches of our government functions.[103] It wouldn't even make a dent in the monthly auction of Treasury debt, and it would allow the Judiciary to free PACER and lose nothing.[104]

14.3.2 *Infrastructure Improvements*

As I discuss above, one of the real problems with PACER is that it provides very limited search capacity. The House bill introduced in the last Congress had modernization features directed at this problem,[105] and the Senate bill spoke

[100] *See, e.g.*, Judiciary Info. Tech. Fund (providing 2020 estimate of $147.7 million in "Estimated Receipts and Prior Year Recoveries" for the revenue on "EPA Program," which includes PACER, in Table 11.1 at page 11.2).

[101] U.S. Fish & Wildlife Serv., 2016 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation 90 tbl.39 (2016), https://www.census.gov/content/dam/Census/library/publications/2018/demo/fhw16-nat.pdf (reporting $4 billion in such spending in 2016).

[102] *Federal Court Funding*, A.B.A. (Dec. 31, 2020), https://www.americanbar.org/advocacy/governmental_legislative_work/priorities_policy/independence_of_the_judiciary/federal-court-funding/.

[103] *See also* Sanga & Schwartz, *Tear Down This Judicial Paywall* (proposing abolition of PACER fees and advocating that "[t]o cover the shortfall, Congress should increase judicial appropriations accordingly").

[104] If Congress won't do that, the Judiciary should find other ways to increase access. It could, for example, eliminate fees for all documents and information docketed more than – say – a year earlier. That would probably preserve the Judiciary's ability to charge substantial download fees, assuming that the institutional customers that pay most of the fees need data quickly. And the operating cost to the Judiciary of ensuring that PACER's contents are available to the public for useful research is easy to approximate: zero dollars and zero cents. That's true because the Internet Archive has already offered to host the data for free; *see* Caitlin Olson, *Internet Archive Offers to Host PACER Data*, Internet Archive Blogs (Feb. 13, 2017), http://blog.archive.org/2017/02/13/internet-archive-offers-to-host-pacer-data/ (printing Feb. 10, 2017 letter from Brewster Kahle making this offer).

One can think of other smaller-loaf alternatives. For example, the Judiciary could release one or more random samples of documents, perhaps using case-level sampling to allow case-level study. Or it could provide restricted-use access with an approval process, as some other federal agencies do.

[105] H.R. 8235, 116th Cong. § 2(b) (2020).

generally of improvements in "data accessibility … and performance."[106] Not knowing anything about PACER's back-end design and capabilities, I don't know whether its lack of a docket-entry or document text search function is a bug or is deliberate. But there are surely many nonprofit organizations that would and could host the data. As noted, the Internet Archive has already offered to do so for free, and to do so with considerably better functionality than PACER currently has.

There are other infrastructure improvements that the Judiciary could make. Some require more burden than others. Less demanding would be more refined docketing, for example, allowing docket coding for motions to dismiss according to the basis for the motion. On the high end would be redesigned forms such as the civil cover sheet, requiring more and/or better information from litigants.

A group of professors recently discussed many of the pathologies I have already described, in *Science* magazine.[107] They advocate a "collaborative research agenda to empower the public to access and analyze court records."[108] This agenda includes better organization of the data because of the fact that court records tend to include unstructured text. The professors advocate developing

> applications that not only support scholars and researchers who may want to analyze the data but also enable members of the judiciary, entrepreneurs, journalists, potential litigants, and concerned citizens to learn more about the functioning of the courts. To support inquiries made by the public, we should develop applications that can process natural language queries such as "What are the most recent data privacy cases?" or "How often do police officers invoke qualified immunity?"[109]

So there are surely a range of other infrastructure improvements worth considering. My purpose here is only to suggest that more about PACER than just the paywall can be improved.

14.3.3 *The Judiciary Should Adopt an Attitude of Openness*

Certainly there are many judges who embrace openness. Judges have approved many, perhaps even most, PACER fee waivers that scholars have actually submitted to the courts. An amicus brief by retired judges in the ongoing PACER class-action litigation indicates laudable support for public access to court data.[110] Over the years, I have had numerous conversations with professionals at the FJC and AO that indicate a range of support between partial and complete for providing more data access.

[106] S. 2064, 116th Cong. § 3(b)(1) (2019).
[107] Pah et al., *How to Build a More Open Justice System*.
[108] *Id.* at 135.
[109] *Id.* at 136.
[110] Brief of Retired Federal Judges as Amici Curiae in Support of Neither Party, Nat'l Veterans Legal Servs. Program v. United States, No. 19-1081 (Fed. Cir. Jan. 23, 2019), ECF No. 26.

At the same time, it is hard not to read a message of opposition from some of the Judiciary's choices, including setting PACER fees at an unnecessarily high level, opposition to legislation that would increase PACER access (including by circulating boilerplate "talking points" for federal judges to use in lobbying members of Congress),[111] and anecdotal evidence that some judges may reject openness as a categorical matter.[112]

My own experience is at least partially consistent with that picture, and it suggests that this rejection exists not only among particular judges, but also as an institutional fact. In 2015 I attended two workshops about access to federal court data. The first was one I organized at Penn Law, which was funded by the National Science Foundation,[113] and the second was the Federal Courts Civil Data Project Roundtable, which was held in Washington, DC, and hosted by the ABA Standing Committee on the American Judicial System and the Duke Law Center for Judicial Studies. Participants at one or both of these workshops included federal judges; academics who work with federal data; staff researchers and/or attorneys from federal agencies including the Federal Judicial Center, the AO, the Administrative Conference of the United States; a representative from the National Center for State Courts; and Michael Lissner, executive director of the Free Law Project. I have limited space to describe my impressions of the views of the Judiciary's representatives at these events. There were a variety of views, but it seems clear that AO representatives harbor skepticism about the economic, bureaucratic, and technological feasibility of opening up PACER (though the multi-court fee waiver exemption process that now exists occurred not too long after these events).[114]

In addition, at the ABA event, one Article III judge spoke forcefully against additional PACER access. This judge, whom I know from prior communications to be a devoted public servant, pointed to the example of a news story about a colleague that the judge thought was both unfair and professionally embarrassing. The judge at the conference expressed sincere and profound concern that the judiciary would be unduly pressured in this and other ways by the public availability of additional data that could be searched and filtered easily.

[111] *See* John Eggerton, *Fix the Court: Judges Are Lobbying against PACER Reform Bill*, Multichannel News (Dec. 1, 2020), https://www.nexttv.com/news/fix-the-court-judges-are-lobbying-against-pacer-reform-bill.

[112] *See, e.g.*, Engstrom, *The* Twiqbal *Puzzle*; text accompanying note 75 (discussing *In re Application for Exemption from Electronic Public Access Fees for Chantalle R. Forgues, Daniel Lee, and Plymouth State University*).

[113] Increasing Access to Federal Court Data Workshop.

[114] For a somewhat longer discussion, see Jonah B. Gelbach, *Comments about Two 2015 Workshops on Federal Court Data*, JonahGelbach.com (May 5, 2021, 5:00 PM EST), https://www.jonahgelbach.com/blog#h.6fappsisbb0l. Interested readers may also find a discussion of the unfortunate judicial response to my attempts to obtain nationwide data about the demographic composition of jury pools. Jonah B. Gelbach, *Comments about Attempts to Obtain Data from District Court AO12 Forms*, JonahGelbach.com (May 10, 2021, 2:39 PM EST), https://www.jonahgelbach.com/blog#h.1xufllwzpd24.

There is some evidence that the Civil Justice Reform Act distorts how judges approach their dockets, perhaps to avoid embarrassment.[115] I'm an economist who believes that people generally respond to incentives, so I'm not surprised. Even so, Article III judges have substantial constitutional protections precisely to allow them to buck public pressure and embarrassment in the service of judicial independence. Most obviously, they have life tenure, and their salaries cannot be reduced.[116] Their jobs carry considerable prestige, and no small amount of power; Chief Justice Stone once remarked, "the only protection against unwise decisions, and even judicial usurpation, is careful scrutiny of their action and fearless comment upon it."[117] I oppose unfair treatment of judges as much as anyone else, but it should go without saying that unfair treatment is not the same as informed public scrutiny. Citizens should have confidence that the federal judiciary will not be deterred from doing its job with integrity by the availability of such scrutiny, and for this reason the Judiciary should welcome rather than avoid it.[118]

Judges play an important and powerful role in our democracy, and citizens have a right to the fruits of studying how the courts do their, and thus our, business. Any defensible vision for the future of public access to court records should include a decision by the Judiciary to switch gears and foster an institutional attitude of openness toward the public.

14.4 CONCLUSION

There's a lot of federal court data. It sits on servers in district courts around the country, with a single access point through PACER. It's often inconvenient or impossible to use it to study the federal litigation system. Even where PACER use is technically feasible, it's often prohibitively costly in financial terms to acquire the magnitude of data that would allow scholars to do research that would then be available to the courts and the public for free. PACER has severely limited search functionality, and individual courts decide whether even that will be open to

[115] *See* Miguel F. P. de Figueiredo, Alexandra D. Lahav & Peter Siegelman, *The Six-Month List and the Unintended Consequences of Judicial Accountability*, 105 Cornell L. Rev. 363, 364 (2020); Jonathan Petkun, Can (and Should) Judges Be Shamed? Evidence from the "Six-Month List" (Mar. 2020) (unpublished manuscript), https://jbpetkun.github.io/pages/working_papers/SixMonthList_WorkingDraft_20200327.pdf.

[116] Bankruptcy judges do not have such protections. But the private service "AACER" – Automatic Access to Court Electronic Records – has already opened bankruptcy data to public bulk searches. *See Epiq AACER*, Epiq Bankr., https://www.aacer.com/. AACER is not free to the public, however.

[117] Schultze, *The Price of Ignorance* (quoting Alpheus Thomas Mason, *Harlan Fiske* Stone: Pillar of The Law 398 (1956)).

[118] Things could be worse, to be sure. In 2019, France prohibited *all* data analytics using information about judges. *Understanding the French Ban on Judicial Analytics*, Gregory Bufithis (June 6, 2019), http://www.gregorybufithis.com/2019/06/09/understanding-the-french-ban-on-judicial-analytics/.

researchers without their paying ruinous fees to download documents that are owned by the public. All the while, the Internet Archive and who knows how many other nonprofits would gladly crawl, index, and host PACER's data for free public download.

This situation is an intolerable disgrace. It is unacceptable for a democracy to have but withhold such valuable information about the functioning of its courts. With only an imperceptible impact on the public fisc, Congress could both mandate free PACER and preserve the Judiciary's budget.

There do not seem to be large hurdles related to privacy or pressure on judges, because the data in question are already freely available for public download – just at a deliberately high price that renders useful research infeasible. Nor is there any real technical feasibility problem, as the offer by the Internet Archive to host the data for free forever shows. This is a problem of political and attitudinal will. Congress and the Judiciary should solve it, so we can use court data to increase access to better justice.