Testimony of
Thomas R. Bruce

Co-Founder and Director of the Legal Information Institute
A Research and Publication Activity of the Cornell Law School

before the
House Committee on the
Judiciary's Subcommittee on Courts,
Intellectual Property and the Internet

on
Judicial Transparency and Ethics

February 14, 2017

Comments of the Cornell Legal Information Institute
before the
House Judiciary Committee
Subcommittee on Courts and Intellectual Property

February 14, 2017

Chairman Issa, Ranking Member Nadler, and members of the Committee, thank you for inviting me to appear before you today.

My name is Tom Bruce, and I am the co-founder and Director of the Legal Information Institute, a research, engineering, and publishing activity of the Cornell Law School. In 1992, we were the first to make judicial opinions available via the Web; our publication of the decisions of the Supreme Court anticipated the development of their Web site by 8 years. In the intervening quarter-century, we have gained a great deal of expertise in the creation of advanced technologies for legal publishing, some of it in collaboration with groups well-known to this Committee. We have undertaken joint studies with the Government Publication Office, and consulted on advanced legislative metadata models for the Library of Congress. We served on the House Bulk Data Task Force and as members of the steering committee for its annual Legislative Data Transparency Workshop. Perhaps surprisingly, we have never published materials that are taken from PACER, though we are well-acquainted with its use by others. Last year, the LII's web site at law.cornell.edu provided legal information to more than 32 million unique individuals.

I am here to speak to you about only one of the several important matters before the Committee today, namely the operation and future direction of the PACER system for public access to the opinions of the Federal courts. This is not the first time that this matter has come before the committee, and I intend to be brief. I will not revisit the many criticisms of the capabilities of that system, or of its fee structure, beyond the bare minimum necessary to get a glimpse of a useful way forward.

What PACER is

With that in mind, let me focus on three things that define PACER:

First, PACER charges fees for access to public records. That has been the cause of a great deal of criticism[1], not only because fees erect a barrier for many, but because the revenue from fees at current levels considerably exceeds the cost of operating the system. The excess revenue is diverted for use on

[1] Many have taken issue with the charging of fees, and particularly fees beyond cost recovery. See, for example, https://blog.law.cornell.edu/voxpop/2011/02/03/pacer-recap-and-the-movement-to-free-american-case-law/, for an article by Steven Schultze describing the inception of the RECAP project. Wikipedia provides a good list of sources in its treatment of the system, at https://en.wikipedia.org/wiki/PACER_(law), including an article from the New York Times that describes PACER as "cumbersome, arcane, and not free". The Free Law Project's "Downloading Important Cases on PACER Costs More Than A Brand New Car" uses a whimsical method to describe the problem in very concrete terms, comparing the cost of PACER research in a major case to the cost of a Honda Civic. See https://free.law/2016/11/17/downloading-important-cases-on-pacer-costs-more-than-a-brand-new-car/

The fee schedule is currently the object of a class-action lawsuit initiated by three non-profit organizations. See Barry, Kyle. "Alliance for Justice sues the Administrative Office of the U.S. Courts for charging excessive and illegal fees to access court records". Alliance for Justice. (at http://www.afj.org/press-room/press-releases/alliance-for-justice-sues-the-administrative-office-of-the-u-s-courts-for-charging-excessive-and-illegal-fees-to-access-court-records) (retrieved February10, 2016).

other projects. That is unjustifiable and inconsistent with the policies established by the Congress in the E-Government Act of 2002. [2]

The fee issue is compelling, but it is not the only issue. Dropping fees altogether would be laudable, in that it would remove the economic barriers to public access. It would not relieve other problems, notably with outdated technology[3] and with usable citation – and indeed it might require that the Administrative Office address long-neglected problems with personally-identifiable information in the database.

Second, PACER became outmoded two years after it was built, and in some ways has never caught up. That was, to some extent, an accident of history. Implemented only two years before the introduction of the World-Wide Web created a revolution of rising expectations for online information systems, it was all too quickly seen as outmoded and out of touch with current technology. Unfortunately, improvements came slowly and the gap widened. Recent progress has been more rapid, but the system still falls short on a number of dimensions, notably in the area of search and retrieval[4].

In 1990, there was very little expertise in the design and operation of large-scale case-management and legal-publishing systems outside the two largest commercial legal-information services (recall that this was a time when the Justice Department was, at high cost, buying back its own work product from what was then the West Publishing Company[5], now a division of the Canadian company Thomson-Reuters, much as the government continues to buy its own work back from PACER today). In 2017 the situation is radically different. Expertise is much more widespread, diffused across multiple companies, non-profits, and academic institutions. Legal-publishing and case-management companies are numerous and there is vigorous competition both for market share and for technological advantage[6] at all price points. That is a market that would get further stimulus from a more open PACER.

Innovative approaches are also flowing from the non-profit sector and from government. The Free Law Project[7], the Internet Archive[8], and the FDsys[9] collection jointly operated by the Government

---

[2] The E-Government Act of 2002 (PL 107-347), section 205(a), provides for public access to the opinions of the Federal courts via website. Section 204(e) amends the Judiciary Appropriations Act to read, " the Judicial Conference may, only to the extent necessary, prescribe reasonable fees… to reimburse expenses incurred in providing these services."

[3] As with the fee schedule, a good list of criticisms of PACER's technology is in its entry in Wikipedia (https://en.wikipedia.org/wiki/PACER_(law)), for example Greg Beato's "Tear Down This Paywall" in the June 2012 issue of *Reason* magazine (at http://reason.com/archives/2012/05/30/tear-down-this-paywall), which describes it as "antiquated as a barrister's wig". More recently, the principals of the Free Law Project have published a series of essays on the problems with the system, beginning at https://free.law/2015/03/20/what-is-the-pacer-problem/.

[4] see note iii above.

[5] See, *eg.*, Wolf, Gary, "Who Owns the Law?", WIRED Magazine issue 2.05, May 1994, online at http://archive.wired.com/wired/archive/2.05/the.law.html .

[6] There has never been more robust activity in legal technology than there is at present. The current environment is rich in startup activity; for examples, see documentation on the "Reinvent Law" events held in New York, London, Dubai, and Silicon Valley (many of the talks are available at www.reinventlawchannel.com). At the time of writing, Robert Ambrogi's list legal tech startups numbered 614. It is particularly detailed and helpful. See www.lawsitesblog.com/legal-tech-startups .

[7] https://free.law . The site offers millions of opinions from 420 jurisdictions (see https://www.courtlistener.com/coverage/), and is the current home of the RECAP PACER-harvesting project.

[8] The Internet Archive, at https://archive.org, provides archives of digital documents and multimedia materials at

Publication Office and the Administrative Office of the Courts all offer capabilities that exceed those of PACER in significant ways, particularly in the area of full-text search of cases and its integration with available metadata. Similar examples exist among the state courts, notably at the site maintained by Ohio's Reporter of Decisions for that state's Supreme Court[10], and at the opinions archive run by the Illinois Reporter of Decisions[11].

Third, PACER suffers from a split personality. On one hand, it is an electronic filing and case management system that supports the Federal courts, with an audience of lawyers, judges, and court administrative personnel. On the other – and most important to the public – it is a data-publishing system that offers the work of the Federal courts, both documents and metadata, to a very wide range of people, including litigants, researchers, and government bodies outside the judiciary. It is on PACER's data-publishing function that I will focus now.

That split personality is very much on display in a 2015 article[12] written by two senior staffers from the Administrative Office of the US Courts. Each of the authors has been involved with the design and management of the PACER system for more than 38 years. Their article describes the creation of the specifications for the "NextGen" PACER system. All but a very few of the improvements described in the article are aimed at the e-filing and case-management side of PACER. The innovations intended for the public were largely confined to streamlining the process by which they might file for bankruptcy.

What PACER is not

Equally, there are a number of things that PACER is not.

It is not transparent in its business model or operations. In preparing this testimony, I was repeatedly struck by the difficulty of acquiring information about the design and operation of the system, and about details of the business model on which it is based. I was fortunate to be able to draw on the work of academics and others who have devoted considerable time to puzzling out the little that is known to outsiders.[13] We should all be grateful for their work.

PACER is not an adequate facility for research on the activities of the Federal courts. Social scientists, legal scholars, linguists, and administrators who want to increase the efficiency of court activities – indeed, researchers in a great many disciplines – do not have useful access to PACER's data. That is chiefly because it does not provide bulk access to that data. Significantly, research activities that might be carried out on behalf of the Congress itself are equally impeded. Social Security cases, prisoner

---

staggering scale. A look at https://archive.org/search.php?query=judicial+opinions reveals that the collection is rich in judicial documents and commentary, including but by no means limited to an archive of the RECAP project.

[9] FDsys has its main page at https://www.gpo.gov/fdsys/

[10] The State of Ohio Supreme Court site is at http://www.supremecourt.ohio.gov/.

[11] The Illinois opinions archive is at http://www.illinoiscourts.gov/Opinions/archive.asp.

[12] Brinkema, John, and J. Michael Greenwood. *E-Filing Case Management Services in the US Federal Courts: The Next Generation: A Case Study*. International Journal for Court Administration, vol. 7, n. 1, July 2015. URN:NBN:NL:UI:10-1-115635 . Available online at http://www.iacajournal.org/articles/10.18352/ijca.179/galley/191/download/.

[13] I am especially grateful to Steven Schulze and Carl Malamud, who provided me with comprehensive lists of their earlier work on this subject.

appeals, and immigration matters are all examples of areas in which study of judicial outcomes is important to those who have responsibilities that call for the investigation and evaluation of operations across the full breadth of our system of government. As a practical matter, most of these problems stem ultimately from PACER's failure to provide access to data in bulk.

PACER is not an effective protector of privacy. It contains, and exposes, any amount of personally-identifiable information useful for identity theft[14]. It does not do a good job of protecting the identities of crime victims or of helpful informants (as witness the existence of the website **whosarat.com**[15]). We cannot know the full extent of these problems because, without bulk data access, research or assessment across the full scope of the database is practically impossible.

PACER is not an adequate vehicle for citable legal research. In particular, it does not provide vendor- and medium-neutral identifiers that could provide a basis for either permanent or interim citation, and it retains pagination as the basis for pinpoint citation[16]. Indeed, any identifier that conformed to a uniform scheme for uniquely identifying the opinions of the Federal courts would provide the basis for connection with more traditional citation schemes, but that is lacking.

These are stubborn problems. The sheer size and scope of the document database, the diversity and lack of uniform editorial and classification standards among the courts that originate the documents, and the sensitivity of some of the information all present daunting challenges, some of which have been capably dealt with.

The remaining challenges are not insuperable, provided that the Congress acts. Both the Congress and the Federal courts have strongly and repeatedly announced their commitment to providing full access, even to unpublished opinions, at minimal or no cost. The sentiments expressed by the Congress in the E-Government Act of 2002[17] are echoed in the statements of Justice Alito's committee report supporting Rule 32.1 of the Federal Rules of Administrative Procedure. That Committee pointed out that the E-Government Act of 2002[18] mandated the federal courts (trial courts as well as appellate) to place all their opinions on public Web sites in a text-searchable format – "regardless of whether such opinions are to be published in the official court reporter."[19] Wrote the committee: "The disparity

---

[14] In 2011, Timothy Lee reported research on redaction failures at https://freedom-to-tinker.com/2011/05/25/studying-frequency-redaction-failures-pacer/ . Carl Malamud has done similar work with the detection of Social Security numbers in a small slice of the opinions available in PACER. Interestingly, James Grimmelmann has pointed out that the removal of paywalls from PACER would, to the degree that economic barriers provide practical obscurity, worsen the privacy problem. See https://arstechnica.com/tech-policy/2009/04/case-against-pacer .

[15] https://www.whosarat.com . Founded by an embittered former DEA informant, the site purports to provide information about "informants and agents", information that it at one time acquired by mining PACER for data on individuals who had plea-bargained in multiple Federal criminal cases.

[16] See generally Martin, Peter W., "One District Court's Lonely Gesture Toward Open Access and Medium-Neutral Citation", in his "Citing Legally" blog at http://citeblog.access-to-law.com/?p=797 .

[17] See note ii, above.

[18] See note ii, above.

[19] Memorandum from Hon. Samuel A. Alito to Hon. David F. Levy, at 4 (May 6, 2005), available at http://www.uscourts.gov/rules/Reports/AP5-2005.pdf . Quoted in Martin, Peter, *Finding and Citing the 'Unimportant' Decisions of the United States Courts of Appeal* (2008), online at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1125484 .

between litigants who are wealthy and those who are not is an unfortunate reality. Undoubtedly, some litigants have better access to unpublished opinions, just as some litigants have better access to published opinions, statutes, law review articles – or, for that matter, lawyers."[20] But the report continued: "[T]he solution is found in measures such as the E-Government Act, which makes unpublished opinions widely available at little or no cost."[21]

What should be done

1. *Fees need to be removed as quickly as possible.* Dissemination fees have strongly inhibited beneficial uses of the data contained in the primary record of the workings of our Federal courts. Consideration should be given to removing per-page viewing fees, or at the very least paring them back to a level that more closely matches PACER's cost of operation. A fee schedule that generates a surplus clearly disregards the will of the Congress as it was stated in the E-Government Act of 2002[22], and subsequently relied upon by Justice Alito's Committee on Appellate Rules in 2005[23]. If instantaneous removal is too disruptive to the processes of judicial administration, a brief sunset period might be considered.
2. *The details of PACER's operations and business model need to be far more visible to the Congress and to the public.* It is nearly impossible for Congress to assess the problem, or for outsiders to make responsible recommendations, given the lack of transparency around PACER. It is far too difficult to find out – for example -- what percentage of PACER's revenue comes from filing fees, how much is derived from data sales to for-profit entities, what expenses are incurred by maintenance and improvements and so on and on. A CRS report describing the business and technical operation of the system in detail would be more than helpful, and would bring welcome clarity to many of the issues involved. To give two examples, outsiders have suggested that the total cost of operation of the system might be recouped exclusively from filing fees, if there were a modest increase; it is also possible that a licensing system that required commercial users of legal information to pay reasonable fees for the raw materials on which their products and businesses are built might do equally well. But without detailed information it is impossible to know for certain, or even to make responsible suggestions.
3. *The users of PACER's data-publication services need representation in the planning and design processes*. The previously-mentioned article on NextGen design[24] shows that input into system design has come exclusively from within the judiciary and from a few "power users" of the e-filing and case-management systems. The designers even chose to ignore the recommendations of their own hired experts – consultants from from MITRE Corporation, a well-respected consulting group that has successfully applied technology to many aspects of judicial administration. There appears to have been little or no consultation with those outside the judiciary who use the publication system, or with outside experts in online dissemination of legal information.
4. *PACER'S data-publishing activities should move to a new home.* The article about PACER's

[20] *Id.*, at 6.

[21] *Id.*

[22] See note ii, above.

[23] See note xix, above.

[24] Brinkema and Greenwood, note xii above, p.4

NextGen design, written by two very senior PACER staff members, celebrates the responsiveness of the NextGen design to the needs of the judiciary and to a small group of users of the filing and case-management systems[25]. That can be charitably interpreted as a sound effort to respond to the range of important customers who were in a position to express their needs to the designers. Understandably, those to whom the designers are most immediately responsible are preoccupied with the e-filing and case-management portions of the system -- indeed, the Federal judiciary has from the earliest times preferred to let others take on the chore of publishing their opinions.

Why not, then, put responsibility for data-publishing activities with an organization that has publishing as its primary mission, and the experience and expertise to successfully engage the challenges that the data-publishing side of PACER presents? The Government Publishing Office already has a pilot program for the publication of judicial opinions underway[26]. It is a joint undertaking with the Administrative Office of the Courts. In many ways, it has been highly successful and appears to be scalable to the dimensions that PACER would require. Obviously, the technical problems of transfer from the PACER system have been worked out to a degree, with success that can be built on.

Some assembly will be required. The FDsys collection is based on sound technical underpinnings and data models, but it would need significant expansion. At the moment it covers a relatively small number of courts, and does not extend to the full chronological range available from many of them[27]. The metadata associated with each document is, by comparison with PACER, woefully incomplete (for example, it does not currently contain dates of decision or the names of opinion authors)[28]. Metadata associated with documents in FDsys would need to be brought up, immediately, to the level of embedded metadata available from commercial systems. Removal of the paywall would increase the need for attention to long-neglected privacy concerns.

Ultimately, for the sake of policy, practicality, and fairness, those outside the judiciary should have the same tools available to them as those within it. And ultimately all public data in PACER should be published in formats that encourage its use, using apparatus that facilitates use in bulk[29].

Conclusion

The benefits of bringing PACER back into line with its Congressional mandate, increasing the transparency of its operations, and of placing its publication activities in the hands of those better

---

[25] Brinkema and Greenwood, note xii above, p.5.

[26] The FDsys USCOURTS collection is very briefly described here: https://www.gpo.gov/help/index.html#about_united_states_courts_opinions.htm

[27] The collection currently contains opinions from 110 courts, representing 885,000 opinions. The collections date back to approximately 2004; there is no set schedule for complete coverage. Private communication from Lisa LaPlant, FDsys program manager, February 7, 2017.

[28] A list of the metadata fields and values available from the FDsys USCOURTS collection is at https://www.gpo.gov/help/index.html#about_united_states_courts_opinions.htm .

[29] The Free Law Project provides bulk data services via their web site at https://free.law . Examples of bulk metadata services for government information abound; the best ones at present are outside the United States, although courts everywhere have been slow to provide access to their metadata in bulk. For example, see the UK Data.gov.uk project at https://data.gov.uk/ .

equipped to carry them out, are many. First and foremost will be the removal of barriers that prevent the public from reading the opinions of the courts for themselves, and from exercising the right to know the laws that govern them. Publication systems that permit research utilizing the full range of data available from PACER will make it easier for the Congress to fulfill its responsibilities, improve the efficiency and functioning of the judiciary, and stimulate new approaches to legal information and encourage new and innovative businesses.

Thank you for the opportunity to testify today. I look forward to your questions.

Notes