

February 10, 2017

The Honorable
Darrell Issa, Chairman

The Honorable
Jerry Nadler, Ranking Member

Subcommittee on Courts, Intellectual Property and the Internet
Committee on the Judiciary
House of Representatives
Washington, DC 20515

Dear Chairman Issa and Ranking Member Nadler,

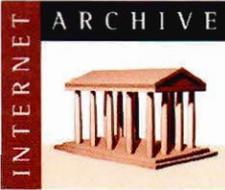
Thank you for the opportunity to submit comments on the Judiciary Committee's hearing entitled "Judicial Transparency and Ethics." I write on behalf of the Internet Archive, a non-profit digital library that is based in San Francisco with facilities throughout the world.

For more than 20 years, the Internet Archive has been archiving digital collections and making them available at no cost and with no restriction on the Internet. The Internet Archive works with the Library of Congress, the National Archives, and numerous national libraries around the world to collect, store, and provide permanent access to millions of books, videos, audio and hundreds of millions of pages of U.S. government documents, including over 14,000 hours of video of Congressional hearings.

By this submission, the Internet Archive would like to clearly state to the Judiciary Committee, as well as to the Administrative Office of the U.S. Courts and the Judicial Conference of the United States, that we would be delighted to archive and host—for free, forever, and without restriction on access to the public—all records contained in PACER.

People download more than 20 million books from the Internet Archive each month. We preserve 1 billion web pages each week for public access through the "Wayback Machine." Indeed, the Wayback Machine is the only publicly accessible archive of all the web sites of Congress. At any given moment, we are delivering about 30 gigabits of data per second. We host more than 20 petabytes of data in total.

By comparison, the PACER corpus is a fraction of a petabyte and does not use a significant amount of bandwidth. We have the capacity to host this information, and I know there are many other organizations on the Internet who would be able to make dramatic increases in the usability and utility of our Federal Judiciary's database if it were made available in a more modern fashion and without artificial restrictions on use.



The stated purpose of PACER is to make public court records “freely available to the greatest extent possible.” Sixteen years ago, the United States Courts predicted that PACER would allow the public to “surf to the courthouse door on the Internet.” Today, anyone visiting a federal courthouse can view the public record for free. PACER, on the other hand, charges users per-page fees that are prohibitive for many members of the public. The Judiciary could resolve this unfortunate discrepancy—immediately—at no cost. This is our offer.

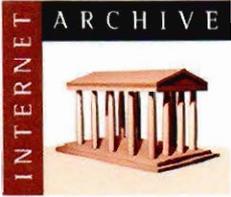
The Internet Archive has deep experience with collections of this kind. In fact, we already host the records from over a million federal court cases that have been donated by the public as part of the RECAP Project. However, 1.4 million is a small portion of the hundreds of millions of cases that PACER contains, and we are frustrated that it is so difficult to obtain and serve the workings of our federal courts to the public. This is fairly trivial technical task, and we would welcome the opportunity to make much more data available.

I must also note that The Internet Archive is not alone in being well-equipped to offer this service. There are other large digital repositories that similarly serve the public for free. I cannot speak for them, but I believe that once the corpus is available for no fee and without restriction, they too will replicate it and offer similar service. Indeed, others may build useful tools for reading, searching, and studying the corpus of public court records that makes up our federal case law.

In order to recognize the vision of universal free access to public court records, the Federal Judiciary would essentially have to do nothing. We are experts at “crawling” online databases in an efficient and careful fashion that does not burden those systems. We are already able to comprehensively crawl PACER from a technical perspective, but the resulting fees would be astronomical. The Federal Judiciary has a Memorandum of Understanding with both the Executive Office for US Trustees and with the Government Printing Office that gives each entity no-fee access for the public benefit. The collection we would provide to the public would be far more comprehensive than the GPO’s current court opinion program—although I must laud that program for providing a digitally-authenticated collection of many opinions.

By making federal judicial dockets available in this manner, the Federal Judiciary would enable free and unlimited public access to all records that exist in PACER, finally living up to the name of the program. In today’s world, public access means *access on the Internet*. Public access also means that people can work with big data without having to pass a cash register for each document.

This PACER collection we would maintain and improve would have far more detailed metadata and contextual information than the GPO service or the PACER Case Locator service. And, that’s just for starters, because we know that there are thousands of eager researchers, journalists, and government workers



(including Congressional staff) who would immediately jump in and work with us.

By providing no-cost access to the Internet Archive to PACER and accepting our commitment to make this information available for use without restriction in perpetuity, we believe we can work with our government to make the workings of our court more usable to government attorneys, to members of the bar, and to the public at large.

Sincerely yours,

A handwritten signature in blue ink, appearing to be "Brewster Kahle".

Brewster Kahle
Digital Librarian and Founder, Internet Archive