



PUBLIC.RESOURCE.ORG ~ A Nonprofit Corporation

Public Works Projects for the Internet

October 3, 2008

08-AP-B

The Honorable Lee H. Rosenthal
Chair, Committee on Rules of Practice and Procedure
Judicial Conference of the United States
Washington, D.C. 20544

08-CV-D

Dear Judge Rosenthal:

I would like to thank you for your letter of July 16, 2008 on the subject of personal identifiers in appellate opinions. Your kind words are very much appreciated and I am pleased to report that the Clerks of the Courts of the Second, Fifth, and Ninth Circuits wrote to me indicating they were in the process of redacting social security numbers.

One issue in regards to appellate opinions that I would like to bring to your attention is the status of Alien Identification Numbers. It is the position of the Clerks of the Courts that Alien Identification Numbers do not fall within the enumerated list of "individuals' Social Security and taxpayer identification numbers, names of minor children, financial account numbers, dates of birth, and, in criminal cases, home addresses." I do understand that a literal reading of the list might preclude Alien Identification Numbers and thus bring it to your attention in case the issue had not been previously considered.

I am also writing to you today to report on preliminary results of an audit of documents submitted to the United States District Courts. A social security number scan of these documents shows approximately 2,282 suspect documents in 32 different districts. The social security numbers are present in documents filed in earlier years, but also in many documents filed in 2008. In some cases, it appears that the social security numbers for attorneys and state employees are being disclosed.

While most documents contain the social security number for a single individual, we have found lists of dozens of individuals. In some cases, the name, date of birth, social security number, and even financial account numbers are present, making this "one-stop shopping" for potential identity theft.

I have enclosed for your reference a DVD of the 2,282 suspect documents. You will find attached to this letter as Appendix A a detailed analysis of 13 of the District Courts based a systematic manual scan of the documents flagged by our program. We will be completing the same detailed analysis of the remaining 19 districts for which we have data, and would be happy to forward that information to you if you wish.

It is worth mentioning that the number of privacy incidents varies widely by district. For example, we were unable to find any social security numbers for the Southern District of Texas or the District of Oregon, and the District of Minnesota had only 6 cases with problems, all from 2005 and 2006.

After working with government data for two decades, I am always impressed by the impact the Internet has on the dissemination of public data. The process of learning how to disseminate public databases effectively is one of trial and error and of progressively perfecting the process. The rules and procedures to protect personal identifiers developed by the Committee on Rules of Practice and Procedure are, I believe, a very important step in this regard.

Based on our experience with scanning District Court documents, I hope you will permit me to offer three suggestions that might provide additional support to the goal of broad dissemination of public information while protecting the privacy of individuals.

First, there is no obvious way for a member of the public or a nonprofit research group such as ours to alert the Administrative Office of the Courts to privacy issues. No system is perfect, and the feedback from users of the system is an essential step in finding mistakes before they spread. Many organizations have found that appointing a Chief Privacy Officer provides a single point of contact for the public.

Second, when problems are found, there does not appear to be a systematic way of alerting the providers of legal information. Even though the social security numbers from appellate opinions were removed from court web sites, they are still present on West Law and Lexis Nexis. A notification mechanism when cases are withdrawn or changed would be extremely useful. Such a system should go beyond the commercial services to include the large number of nonprofit groups that disseminate the law. Our own computers at Public.Resource.Org, for example, serve 1 million unique visitors per month, and that number is far larger when we include other sites that copy our data.

Third, while the first line of defense for protection of privacy is with the lawyers who file documents in the PACER system, we must assume that no system is perfect. I have attached as Appendix B a simple one-line PERL program based on open source tools which we use to scan for social security numbers. We scan a database for potential hits and then look at each case manually. If we find a social security number, we use redaction tools to remove that information.

There are no doubt far more sophisticated tools available, but I offer this simple mechanism as an example and would be more than happy to discuss these tools with technical staff if that is useful.

Thank you again for your responsiveness and quick action on the matter of Appellate decisions. It is gratifying to see the commitment towards the protection of personal privacy, both in the Judicial Conference and in the day-to-day operations of the Clerks of the Court.

Very truly yours,



Carl Malamud
President & CEO
Public.Resource.Org

cc: Mr. Peter McCabe, Esq.
The Honorable James C. Duff