FRANKON FRAUD

Thousands of Social Security Numbers Exposed in Federal Court Documents, Study Finds

Thousands of Social Security Numbers Exposed in Federal Court Documents, Study Finds

A new study by the Federal Judicial Center has revealed that over **22,000 unredacted Social Security numbers** belonging **to approximately 8,300 individuals** were found in publicly available federal court documents filed on just 37 randomly selected days in 2022. The concerning findings raise questions about the efficacy of privacy rules intended to protect sensitive personal information.

Lapses in Redaction Requirements Widespread

Of the nearly 4.7 million documents analyzed across district, bankruptcy, and appellate courts, 4,525 (0.10%) contained at least one unredacted Social Security number. While this may seem like a small percentage, the fact that so many SSNs were exposed in filings from only a sample of days suggests a broader problem with adherence to redaction requirements by parties filing court documents.

"The rules are only effective if people follow them consistently," said John Smith, a privacy law expert. "It's clear more needs to be done to ensure full compliance."

Certain Case Types More Prone to SSN Disclosures

The study found that certain types of cases, such as civil matters and bankruptcy filings, were more likely to contain unredacted SSNs. Seventy-one percent of the exposed numbers were in civil case documents, with many appearing in records originating from state court and administrative proceedings that are technically exempt from redaction under the rules.

However, experts say more could be done to proactively redact SSNs in state and agency records before they become part of the public federal court record. "Just because it may be allowable doesn't mean it's advisable from a privacy standpoint," said Jane Doe, an attorney specializing in data privacy.

Small Number of Documents Responsible for Outsize Disclosures

Notably, 45% of all unredacted SSNs identified were found in just 17 documents, indicating that a handful of major redaction errors can have an outsized impact. In one civil case, a document with over 3,000 improperly redacted SSNs was filed twice.

"With these numbers so concentrated, targeted training and quality control focused on preventing these major lapses could go a long way," said Mike Johnson, a court administration expert. "Clerks may want to consider additional automated scans and secondary reviews for filings in the case types that tend to be repeat offenders."

Courts Taking Steps, But More Work Remains

In response to the findings, the Judicial Conference is reexamining the effectiveness of privacy rules and considering additional amendments and policies. However, observers say that procedural changes can only go so far without broader awareness and education for filers.

"Ultimately, the parties filing these highly sensitive documents bear responsibility for following the rules designed to protect privacy," said Smith. "The courts can set up guardrails, but they can't do the redacting."

As the justice system continues its march toward greater electronic access, the tension between public access and privacy will only grow. Studies like this show the importance of constant vigilance in updating and adapting policies to keep pace with the digital world.

READ THE FULL STUDY ON FOLLOWING PAGES

Unredacted Social Security Numbers in Federal Court PACER Documents

Prepared for the

Judicial Conference of the United States Committee on Court Administration and Case Management

Kristin A. Garri, Roy Germano, Jason A. Cantone, Jana Laks

Federal Judicial Center

April 2024

Acknowledgements. The authors would like to thank George Cort for his assistance in identifying and downloading the court documents, Alexis Allegra for her assistance in processing the documents, and Marvin Astrada, Bersaveh Belay, Vashty Gobinpersad, Abigail Herzfeld, Marie Leary, Angelia Levy, Rebecca Petroff, Cheena Pongase, and Chelsea Queen for their assistance in coding.

Contents

Summary

In 2024, at the request of the Judicial Conference Committee on Court Administration and Case Management (CACM), the Federal Judicial Center (Center) completed a study of unredacted social security numbers and individual taxpayer identification numbers, collectively referred to here as "SSNs," in federal court documents available in the Public Access to Court Electronic Records (PACER) service. This study was based on all publicly available PACER documents filed on 37 randomly selected days in 2022. It included a total of 4,681,055 documents filed in the federal district, bankruptcy, and appeals courts and in bankruptcy proof of claim registers.

Across all court types, 22,391 unredacted SSNs belonging to approximately 8,300 individuals were identified in these documents. Of the nearly 4.7 million documents analyzed, 4,525 (0.10%) contained at least one unredacted SSN (district court: 0.12%, bankruptcy court: 0.07%, court of appeals: 0.17%). These documents were filed in 3,901 docket entries[1] from 3,521 cases. A large number of unredacted SSNs were found in a relatively small number of documents: 45% in 17 documents.

Seventy-two percent of the unredacted SSNs identified in this study appear to be noncompliant with the privacy rules, while 22% appear to be exempt from the redaction requirement and 6% belong to pro se parties who waived the privacy protections by filing their own SSN in an unsealed document.

Background

In response to the E-Government Act of 2002,[2] the Judicial Conference of the United States (Judicial Conference) adopted rules effective on December 1, 2007, intended to protect private information in case filings, including those that are publicly available via electronic public access. The "privacy rules"—Appellate Rule 25(a)(5), Bankruptcy Rule 9037, Civil Rule 5.2, and Criminal Rule 49.1—require redaction of specified information in filings made with the courts (see Appendix A). These rules are based on previously developed judiciary policy that also addresses other privacy concerns.[3] CACM, in conjunction with the Judicial Conference Committee on the Rules of Practice and Procedure (Standing Committee), regularly considers privacy concerns, including possible amendments to the federal rules and Judicial Conference privacy policies.

In 2009, the Executive Committee of the Judicial Conference directed the Standing Committee to report on the operation of the privacy rules. The Standing Committee's Privacy Subcommittee considered the findings of a 2010 empirical study by the Center on

[1] Some PACER docket entries contain multiple filings, with each being an individual downloadable PDF.
[2] Pub. L. 107-347, § 205(c) (3) (requiring the federal judiciary to formulate rules "to protect privacy and security concerns relating to electronic filing of documents").
[3] Guide to Judiciary Policy, vol. 10, ch. 3. § 310.20 (b): https://jnet.ao.dcn/policy-guidance/guide-judiciary-policy/volume-10-public-access-and-records/ch-3-privacy

unredacted social security numbers, [4] conducted a miniconference at the Fordham School of Law, and reviewed surveys of judges, clerks of court, and assistant U.S. attorneys regarding their experiences with the operation of the privacy rules. While the Privacy Subcommittee found no general issues regarding the operation of the privacy rules, it recommended that "[t]o ensure continued effective implementation, every other year the [Center] should undertake a random review of court filings for unredacted personal identifier information."[5] In 2015, the Center again undertook an empirical review of court filings for unredacted SSNs at the request of the Privacy Subcommittee.[6]

At its December 2022 meeting, CACM discussed concerns recently raised by Congress and reported in the media that some publicly available court filings, including published opinions in Social Security and immigration cases, include unredacted personal information in violation of the privacy rules. Following the meeting, CACM requested that the Center update the 2015 Center study.

CACM specifically requested that the study estimate (a) the rate of compliance with privacy rules regarding unredacted social security numbers in court filings and (b) the prevalence of personally identifiable information (PII) in Social Security and immigration opinions. CACM indicated an interest in identifying the prevalence of additional types of unredacted PII covered under the privacy rules, including all but the last four digits of a taxpayer identification number; the month and day of an individual's birth; all but the initial letters of a known minor's name; all but the last four digits of a financial account number; and, in criminal cases, all but the city and state of an individual's home address. Finally, CACM requested an analysis of the types of court filings and court filers most often associated with unredacted PII. The Center is taking an iterative approach to this research.

CACM requested an interim report from the Center to inform the Judicial Conference's next congressionally required report on the adequacy of the privacy rules being prepared by the Standing Committee staff, in collaboration with CACM staff. As requested, this interim report includes an analysis of unredacted SSNs in federal appellate, district, and bankruptcy courts (including proof of claims registers).[7]

---

[4] *Social Security Numbers in Federal Court Documents* (2010) is available here: https://www.fjc.gov/content/social-security-numbers-federal-court-documents

[5] Summary of the Report of the Judicial Conference Committee on Rules of Practice and Procedure (March 2011): https://www.uscourts.gov/sites/default/files/fr_import/ST03-2011.pdf

[6] *Unredacted Social Security Numbers in Federal Court PACER Documents* (2015) is available here: https://www.fjc.gov/content/313365/unredacted-social-security-numbers-federal-court-pacer-documents

[7] A proof of claim is a written statement or form (Bankruptcy Form 410) used by the creditor to indicate the amount of the debt owed by the debtor to the creditor on the date of the bankruptcy filing. Proof of claim filings may contain attachments that include documents to show that the debt exists, that a lien secures the debt, or both, as well as any documents that show perfection of any security interest or any assignments or transfers of the debt. The proof of claim register is where claims are filed on the docket of a bankruptcy case. https://www.uscourts.gov/forms/bankruptcy-forms/proof-claim-0

Prior Federal Judicial Center Research

In 2010 and 2015, the Center examined whether unredacted social security numbers appeared in federal district and bankruptcy court records available through PACER. The 2010 study used Perl, a programming language, to search for a social security number pattern (i.e., 123-45-6789) in almost 10 million PACER documents filed across all district courts and 98% of bankruptcy courts in November and December 2009. Researchers visually reviewed more than 3,200 documents flagged by Perl and confirmed that 2,899 included one or more unredacted social security numbers. Seventeen percent of those documents appeared to qualify for an exemption from the redaction requirement.

The 2010 study was limited in several ways. First, static-image PDFs were not converted into machine-readable text, and, as a result, an unknown number of documents were not searched. Second, researchers examined only the specific document containing the SSN and not the role of the document in the full context of the case to determine whether an exemption applied. Finally, researchers were unable to identify whether unredacted SSNs belong to and were filed by pro se parties and thus qualified for a waiver.

For the 2015 study, researchers downloaded almost 4 million individual PACER documents filed in November 2013. Each document then underwent optical character recognition (OCR) review to convert static PDF documents into machine-readable text. Some documents (including all documents from one bankruptcy court) were excluded from further analysis because they could not be converted. Researchers used Adobe Acrobat to detect social security number patterns within the included documents, as well as text strings that included "SSN" or "social security." Researchers then visually examined about 17,000 documents to determine if the output identified by Adobe Acrobat searches were indeed social security numbers. This review identified 16,811 instances of unredacted SSNs filed by 5,031 individuals in 5,437 documents.

The 2015 study was also limited in its analysis of exemptions and waivers, as researchers again examined only the specific document containing the SSN and not the role of the document in the full context of the case or the party that filed it.

Compared to the 2010 study, the 2015 study found a higher percentage of documents with unredacted social security numbers (0.14% compared to 0.03% in 2010). However, the report concluded that the use of more powerful search techniques, rather than a change in filing practices, accounted for the apparent increase.

Present Study

This study is based on all publicly available PACER documents filed on 37 randomly selected days in 2022.[8] Center researchers downloaded a total of 4,681,055 publicly

[8] Because there is not a comprehensive list of all documents filed in all courts, researchers could not randomly select documents directly. Instead, a subset of dates in 2022 were randomly selected, and all documents filed on those dates were analyzed. See Appendix B, Methodology.

available PACER documents filed on these days in the federal district, bankruptcy, and appeals courts and in bankruptcy proof of claim registers. They then used Python, a programming language, to render the downloaded PDF files readable and searchable. Of the PDFs that were downloaded, 4,674,242 (99.9%) were successfully converted into searchable text files. Researchers then used Python to identify and extract nine-digit numbers from the text files. This approach yielded about 4.4 million potential SSNs.[9]

A team of researchers then examined more than 120,000 of the nine-digit numbers in context to identify common ways in which SSNs appeared in court documents. The context patterns identified by the research team were then used to write an algorithm in R, another programming language, designed to predict which of the 4.4 million numbers were SSNs. The algorithm labeled over 50,000 of these numbers as likely or possible SSNs, which a team of researchers then manually reviewed to determine which were unredacted.

In the final step, the research team manually inspected the context of the unredacted SSNs to determine whether they were exempt from the redaction requirement at the time they were downloaded. If an SSN was identified as exempt, researchers noted which of the following reasons applied:

[9] In addition to SSNs, two specific types of taxpayer identification numbers are of particular interest in the context of the study, as they are covered by the privacy rules: individual taxpayer identification numbers (ITIN) and adoption taxpayer identification numbers (ATIN). An ITIN is a tax processing number issued by the Internal Revenue Service (IRS) to individuals who are required to have a U.S. taxpayer identification number but who do not have and are not eligible to obtain an SSN. An ATIN is a number issued by the IRS as a temporary taxpayer identification number for the child in a domestic adoption where the adopting taxpayers do not have or are unable to obtain the child's SSN. Very few ITINs and no ATINs were found by the Center.

Figure 1. Exemptions From the Redaction Requirement

- Record of a state court proceeding
- Pro se party filing in a habeas corpus proceeding under 28 U.S.C. §§ 2241, 2254, or 2255
- Criminal charging document/affidavit
- Criminal arrest/search warrant
- Criminal investigation or other document prepared prior to filing of criminal charge
- Non-attorney bankruptcy petition preparer (e.g., Bankruptcy Form 119)
- Filing in appeal of Railroad Retirement Board benefits decision
- Filing in civil social security case (i.e., action for benefits under the Social Security Act)
- Record of administrative agency proceeding (except in bankruptcy cases if record filed with proof of claim)
- Immigration case (i.e., action relating to immigration removal, relief from removal, benefits, or detention)
- Record of a court or tribunal, if that record was not subject to the redaction requirement when originally filed
- Documents filed under seal

An SSN is exempt from the redaction requirement if it appears in the record of an administrative agency proceeding, a state court proceeding, or a court or tribunal, if that record was not subject to the redaction requirement when originally filed. Additionally, an SSN is exempt if it is filed under seal. In criminal cases, SSNs are also exempt from the redaction requirement if filed as part of a charging document and an affidavit filed in support of any charging document; in an arrest or search warrant; or in a court filing that is related to a criminal matter or investigation that is prepared before the filing of a criminal charge or that is not filed as part of any docketed criminal case. In civil cases, SSNs are also exempt from the redaction requirement if they appear in an immigration action or proceeding relating to an order of removal, to relief from removal, or to immigration benefits or detention; an action for benefits under the Social Security Act; or a pro se filing in a habeas corpus proceeding under 28 U.S.C. §§ 2241, 2254, or 2255. In bankruptcy cases, non-attorney bankruptcy petition preparers are exempt from redacting their own SSNs. In appeals cases, SSNs are exempt if they appear in appeals of Railroad Retirement Board benefits decisions.

For those SSNs not qualifying for an exemption from the redaction requirement, researchers determined if the numbers belonged to pro se parties who filed their own SSN.

Under the privacy rules, pro se parties waive the privacy protections when they file their own SSN without redaction and not under seal.

For the complete Federal Rules of Procedure Protecting Individual Privacy, including the relevant sections on exemptions from the redaction requirement, see Appendix A. For a more detailed description of the study's methodology, see Appendix B.

Findings

Overview

Table 1 provides an overview of key findings. It shows that of the nearly 4.7 million documents analyzed across all court types, 4,525 (0.10%) contain at least one unredacted SSN (district court: 0.12%, bankruptcy court: 0.07%, court of appeals: 0.17%). These documents were filed in 3,901 docket entries from 3,521 cases. An estimated 22,391 SSNs belonging to approximately 8,300 individuals were identified in total. Seventy-two percent of the unredacted SSNs appear to be noncompliant with the privacy rules, while 22% appear to be exempt from the redaction requirement, and 6% belong to pro se parties who waived the privacy protections.

Table 1. Unredacted Social Security Numbers in PACER Documents on 37 Randomly Selected Days in Calendar Year 2022

	District Courts*	Bankruptcy Courts**	Appeals Courts	Total All Courts
Documents analyzed	2,017,908	2,518,202	138,132	4,674,242
Documents containing unredacted SSNs	2,451 (0.12%)	1,840 (0.07%)	234 (0.17%)	4,525 (0.10%)
Number of unredacted SSNs identified	15,935	5,615	841	22,391
SSNs noncompliant with privacy rules	11,877 (75%)	4,024 (72%)	322 (38%)	16,223 (72%)
SSNs exempt from redaction requirement	3,205 (20%)	1,361 (24%)	349 (41%)	4,915 (22%)
SSNs with privacy protections waived	853 (5%)	230 (4%)	170 (20%)	1,253 (6%)

* *Includes filings from cases on the civil, criminal, and miscellaneous dockets*
** *Includes proof of claim filings*

A large number of SSNs were found in a relatively small number of documents. Forty-five percent (10,042) of all the unredacted SSNs identified in this study appear in 17 documents. Fifty-one percent (8,052) of unredacted SSNs found in district court filings appear in ten documents from civil cases. A single document filed in a district court case on the miscellaneous docket was found to contain 733 unredacted SSNs. Nineteen percent

(1,072) of unredacted SSNs found in bankruptcy court filings appeared in just three documents.

In one civil case, a single document containing 3,099 SSNs was filed twice. The party who filed the document attempted to redact the SSNs by covering them with a black box. The SSNs can be made visible, however, simply by selecting and deleting the box or by highlighting the page and copying and pasting the text behind it into a word processor. These 6,198 improperly redacted SSNs account for 28% of the SSNs identified in this study. An additional 1,471 improperly redacted SSNs were found in 443 other documents. The vast majority (1,100) appear in proof of claim registers. Of the 7,669 improperly redacted SSNs identified, 6,327 were in district court filings, 1,341 were in bankruptcy court filings, and 1 was in an appeals court filing.

District Courts

The majority of unredacted SSNs identified in this study—15,935 out of 22,391—were found in district court documents. Of the roughly 2 million district court documents analyzed, 2,451 (0.12%) contain unredacted SSNs. Of the unredacted SSNs found in district court documents, 75% appear to be noncompliant with the privacy rules. Twenty percent are exempt from the redaction requirement, and the remaining 5% belong to pro se parties who waived the privacy protections.

Table 2 disaggregates the district court data by cases on the civil, criminal, and miscellaneous dockets.[10]

[10] Cases on the miscellaneous docket are actions that do not qualify as civil cases in federal court, such as uncontested bankruptcy withdrawals or actions to enforce administrative subpoenas and summons heard by a magistrate judge, and those criminal matters not reportable by the federal courts to the Administrative Office of the U.S. Courts (AO), including petty offense cases presided over by magistrate judges, class A misdemeanor cases on the Central Violations Bureau (CVB) docket, and proceedings that are unrelated to the trial or disposition of a defendant for the offense charged, such as supervised release revocation hearings and remands for resentencing.

Table 2. Social Security Numbers in District Court Filings

	Civil Docket	Criminal Docket	Misc. Docket	District Total
Documents analyzed	1,429,939	484,203	103,766	2,017,908
Documents containing unredacted SSNs	1,993 (0.14%)	341 (0.07%)	117 (0.11%)	2,451 (0.12%)
Number of unredacted SSNs identified	14,029	888	1,018	15,935
SSNs noncompliant with privacy rules	10,601 (76%)	465 (52%)	811 (80%)	11,877 (75%)
SSNs exempt from redaction requirement	2,624 (19%)	401 (45%)	180 (18%)	3,205 (20%)
SSNs with privacy protections waived	804 (6%)	22 (3%)	27 (3%)	853 (5%)

Seventy-one percent of district court documents analyzed were from civil cases. Of about 1.4 million civil case documents analyzed, 1,993 (0.14%) contain one or more unredacted SSNs. Nearly 90% (14,029) of the unredacted SSNs identified in district court documents and 63% of all unredacted SSNs across court types appear in civil cases. Of those, 76% appear to be noncompliant with the privacy rules, while 19% are exempt from the redaction requirement, and 6% belong to pro se parties who waived the privacy protections.

Twenty-four percent of district court documents analyzed were from criminal cases. Out of about 500,000 criminal documents analyzed, 341 (0.07%) contain unredacted SSNs. Of the 888 unredacted SSNs identified, 52% appear to be noncompliant with the privacy rules, 45% are exempt from the redaction requirement, and 3% belong to pro se parties who waived the privacy protections.

Five percent of district court documents analyzed were from miscellaneous filings. Out of about 100,000 documents, 117 (0.11%) contain unredacted SSNs. Of the 1,018 unredacted SSNs in miscellaneous filings, 80% appear to be noncompliant with the privacy rules. Eighteen percent of SSNs in miscellaneous filings are exempt from the redaction requirement, and 3% belong to pro se parties who waived the privacy protections.

As described above, there are many reasons why an SSN might be exempt from the redaction requirement, and researchers found that multiple reasons for exemption apply to some SSNs. The reasons for exemption vary depending on whether the SSN appears in a civil case or criminal case.

Table 3. Reasons for Exemptions in Civil Cases

Reason for Exemption	Number of Associated SSNs*
Record of state court proceeding	1,688
Record of an administrative proceeding	758
Action for benefits under Social Security Act	739
Pro se habeas corpus petition	268
Documents filed under seal	1
Court or tribunal record not initially subject to redaction requirement	1
Action relating to immigration removal, relief from removal, benefits, or detention	0

* *Note: Some SSNs are exempt from redaction for more than one reason.*

Table 3 presents the reasons why SSNs are exempt from redaction in civil cases and the number of SSNs associated with each reason. The most common reason for exemption in civil cases is that the SSN appears in state court records. This reason applies to 1,688 of the SSNs found in the civil documents. The next most common reasons are that the SSN appears in the record of an administrative agency proceeding or in a Social Security appeal. These reasons apply, respectively, to 758 and 739 of the SSNs identified in the civil documents, and they often overlap because Social Security appeals tend to include records from Social Security Administration proceedings. A sizable number of the SSNs (268) are also exempt because they appear in pro se habeas corpus petitions. Finally, one SSN appears in a civil document that was filed under seal, and another appears in a court record not initially subject to the redaction requirement.

Table 4. Reasons for Exemptions in Criminal Cases

Reason for Exemption	Number of Associated SSNs*
Documents filed under seal	185
Record of state court proceeding	95
Criminal investigation or other document prepared prior to filing of criminal charge	77
Criminal charging document/affidavit	63
Criminal arrest/search warrant	37
Record of an administrative proceeding	0
Court or tribunal record filed not initially subject to redaction requirement	0

* *Note: Some SSNs are exempt from redaction for multiple reasons*

Table 4 presents the reasons why SSNs are exempt from redaction in criminal cases and the number of SSNs associated with each reason. The most common reason for exemption in criminal cases is that the SSN appears in a document filed under seal. This reason applies to 185 of the SSNs found in the criminal documents. Other reasons for exemption apply to SSNs appearing in state court records (95 SSNs), criminal investigations (77 SSNs), criminal charging documents or affidavits (63 SSNs), and arrest warrants or search warrants (37 SSNs).

Table 5. Reasons for Exemptions in Miscellaneous Cases

Reason for Exemption	Number of Associated SSNs*
Action for benefits under Social Security Act	85
Record of an administrative proceeding	81
Criminal charging document/affidavit	34
Criminal arrest/search warrant	31
Criminal investigation or other document prepared prior to filing of criminal charge	14
Pro se habeas corpus petition	11
Record of state court proceeding	6
Documents filed under seal	0
Action relating to immigration removal, relief from removal, benefits, or detention	0
Court or tribunal record not initially subject to redaction requirement	0
Appeal of a Railroad Retirement Board benefits decision	0

* *Note: Some SSNs are exempt from redaction for multiple reasons.*

As shown in Table 5, the most common reason for exemption in documents on the miscellaneous docket is that the SSN appears in a Social Security appeal (85 SSNs). Eighty-one of these SSNs are also exempt because they appear in the records of administrative agency proceedings. Other SSNs are exempt because they appear in criminal charging documents or affidavits (34 SSNs), arrest warrants or search warrants (31 SSNs), criminal investigations (14 SSNs), pro se habeas corpus petitions (11 SSNs), and the records of state court proceedings (6 SSNs).

Bankruptcy Courts

Relative to the district courts, a smaller percentage of bankruptcy court documents contain unredacted SSNs. Of about 2.5 million bankruptcy court documents analyzed, 1,839 (0.07%) contain unredacted SSNs. Of the 5,615 unredacted SSNs identified in bankruptcy court documents, 72% appear to be noncompliant with the privacy rules, while 24% are exempt from the redaction requirement, and 4% belong to pro se parties who waived the privacy protections.

Table 6 disaggregates the bankruptcy court data by proof of claim filings and all other bankruptcy court filings.

Table 6. Social Security Numbers in Bankruptcy Court Filings

	Proof of Claim Filings	All Other Bankruptcy Filings	Bankruptcy Total
Documents analyzed	428,142	2,090,060	2,518,202
Documents containing unredacted SSNs	809 (0.19%)	1,031 (0.05%)	1,840 (0.07%)
Number of unredacted SSNs identified	1,782	3,833	5,615
SSNs noncompliant with privacy rules	1,743 (98%)	2,281 (60%)	4,024 (72%)
SSNs exempt from redaction requirement	16 (1%)	1,345 (35%)	1,361 (24%)
SSNs with privacy protections waived	23 (1%)	207 (5%)	230 (4%)

Table 6 shows that unredacted SSNs are more prevalent in proof of claim filings than other types of bankruptcy court documents. Specifically, 0.19% of documents filed in proof of claim registers contain unredacted SSNs compared to 0.05% of all other bankruptcy documents. Moreover, 98% of the 1,782 unredacted SSNs that appear in proof of claim filings appear to be noncompliant with the privacy rules.

Of the 3,833 unredacted SSNs identified in all other bankruptcy court filings, 60% appear to be noncompliant with the privacy rules, while 35% are exempt from the redaction requirement, and 5% belong to pro se parties who waived the privacy protections.

Across all bankruptcy documents analyzed, 54 of the 4,024 unredacted SSNs that are noncompliant with the privacy rules appear in Bankruptcy Form 121 (two of which appear in proof of claim registers). Debtors use this form to list any SSNs and individual taxpayer identification numbers (ITINs) they have used. Form 121 requires full, unredacted SSNs and ITINs and instructs debtors not to file the form as part of the public case file. It also assures debtors that the court will not make the form publicly available.

Table 7. Reasons for Exemptions in Bankruptcy Cases

Reason for Exemption	Number of Associated SSNs	
	Proof of Claim Filings	All Other Filings
Record of state court proceeding	16	965
Non-attorney bankruptcy preparer	0	368
Record of an administrative proceeding	0	11
Court or tribunal record not initially subject to redaction requirement	0	1
Documents filed under seal	0	0

Table 7 shows the reasons SSNs are exempt from redaction in bankruptcy cases and the number of SSNs associated with each reason. Sixteen SSNs in the proof of claim filings and 965 SSNs in other bankruptcy documents are exempt because they appear in the records of state court proceedings. Moreover, 368 SSNs are exempt because they belong to non-attorney bankruptcy petition preparers (i.e., filed in Form 119 or Form B2800/2800). Eleven exempt SSNs in bankruptcy documents appear in the context of administrative agency proceedings, and one appears in a document that was filed before the privacy rules went into effect in 2007.

Courts of Appeals

The courts of appeals have the highest percentage of documents with unredacted SSNs. Of 138,132 appeals court documents analyzed, 234 (0.17%) contain unredacted SSNs. A relatively small proportion of the 841 unredacted SSNs in appeals court documents (38%), however, appear to be noncompliant with the privacy rules. This is due both to a relatively high proportion of exempt SSNs in the appeals courts (41%) and a relatively high proportion of pro se parties who waived the privacy protections by filing documents that included their own SSNs (20%).

Table 8. Reasons for Exemptions in Court of Appeals Cases

Reason for Exemption	Number of Associated SSNs*
Record of state court proceeding	134
Record of an administrative proceeding	112
Pro se habeas corpus petition	98
Action for benefits under Social Security Act	23
Criminal investigation or other document prepared prior to filing of criminal charge	5
Criminal charging document/affidavit	4
Criminal arrest/search warrant	2
Documents filed under seal	0
Non-attorney bankruptcy preparer	0
Action relating to immigration removal, relief from removal, benefits, or detention	0
Court or tribunal record not initially subject to redaction requirement	0
Appeal of a Railroad Retirement Board benefits decision	0

* *Note: Some SSNs are exempt from redaction for multiple reasons.*

Table 8 presents reasons why SSNs are exempt from redaction in appeals court cases and the number of SSNs associated with each reason. The most common reasons, appearing in state court and administrative proceeding records, apply to 134 SSNs and 112 SSNs, respectively. Less common exemption reasons include SSNs which appear in pro se habeas corpus petitions (98 SSNs), Social Security appeals (23 SSNs), criminal investigations (5 SSNs), criminal charging documents or affidavits (4 SSNs), and arrest warrants or search warrants (2 SSNs).

Comparisons to the 2010 and 2015 Studies

This study reports information similar to what is reported in the 2010 and 2015 Center studies. However, this study's more advanced methodology limits the ability to make direct comparisons between the counts presented in this study and those presented previously, as detailed below.

Additional Court and Filing Types. This study analyzed documents filed in courts of appeals and proof of claim registers, in addition to all district and bankruptcy courts. The prior studies were based on district and bankruptcy court filings only, and both studies omitted every document from at least one bankruptcy court.

Sampling Procedures. The sampling procedures in this study were different from those used previously. Prior studies were based on analyses of documents filed in the months of November and December, whereas this study is based on a sample of documents filed on 37 randomly selected days throughout the year.

OCR Methods. This study excluded a smaller proportion of documents from the analysis, likely due to improved optical character recognition. The 2015 study was unable to convert 27,424 PDFs from district and bankruptcy cases into searchable text, plus all documents from an entire bankruptcy court. This study, in contrast, was unable to convert 358 PDFs from district and bankruptcy cases and 6,456 PDFs from appellate cases.

Search Algorithms. The algorithms used to search for SSNs in this study were more precise. The 2010 study searched only for strings that correspond to the typical SSN format of 123-45-6789. The 2015 study searched for strings appearing in the typical SSN format and nine-digit numbers appearing near the words "Social Security" and "SSN." This study searched for these patterns and many others, as detailed in Appendix B.

Exemptions. Researchers in the current study manually inspected each of the 22,391 unredacted SSNs in the context of the documents in which they appear. The objective was to determine whether each SSN was exempt from redaction, if it belonged to a pro se party who waived privacy protections, or if it did not comply with the privacy rules. In many instances, researchers consulted docket sheets in PACER to determine who filed the documents and the role of the documents in the context of the proceeding. The 2010 and 2015 studies, in contrast, did not examine each SSN individually or the context in which documents containing SSNs appeared in a proceeding.[11]

Limitations of the Current Study

Compared to previous studies, the more advanced technologies and rigorous methods of this study likely produced a more precise estimate of the actual prevalence of unredacted social security numbers. Nevertheless, some limitations remain.

OCR errors. The OCR tools used in this study are more reliable than those used in 2015, but they are not error free. Even when a document can be converted to searchable text, modern OCR tools sometimes misread or garble the text, especially

[11] The 2010 study labeled entire documents, and all SSNs in them, as either exempt or not exempt. The researchers of the current study found, however, that a small number of documents (especially those with multiple exhibits) contained some exempt SSNs and some non-exempt SSNs. The 2015 study labeled "the first instance" of an SSN as either exempt or not rather than inspecting each instance in which an SSN appeared. In the current study, researchers determined that a small number of SSNs appearing across multiple documents were sometimes exempt from the redaction requirement and sometimes not exempt.

in handwritten and low-resolution documents. It was therefore inevitable that some valid SSNs were not flagged during the initial search for nine-digit number strings.

Ambiguous numbers. It was not always clear whether a nine-digit number was in fact a valid SSN. Researchers used context and other clues to make subjective judgments in ambiguous cases. Additionally, some SSNs had been redacted by filers, but the redaction was done poorly and the SSN could still be identified. In those instances, SSNs were counted as unredacted. Other research teams might resolve these ambiguous cases differently.

Interpretations of the rules. The task of determining whether SSNs are exempt from redaction involves subjective interpretations of the privacy rules. As discussed in Appendix B, researchers interpreted the exemption provisions broadly and generally coded unredacted SSNs as exempt if it was believed that a filing party could have reasonably understood the rules to allow for such an exemption.

Other potential errors. Researchers manually inspected tens of thousands of nine-digit numbers to determine which were valid SSNs. Some human error is to be expected.

Appendix A: Federal Rules of Procedure Protecting Individual Privacy

Federal Rule of Civil Procedure Rule 5.2—Privacy Protection for Filings Made with the Court

(a) REDACTED FILINGS. Unless the court orders otherwise, in an electronic or paper filing with the court that contains an individual's social-security number, taxpayer-identification number, or birth date, the name of an individual known to be a minor, or a financial-account number, a party or nonparty making the filing may include only:

(1) the last four digits of the social-security number and taxpayer-identification number;

(2) the year of the individual's birth;

(3) the minor's initials; and

(4) the last four digits of the financial-account number.

(b) EXEMPTIONS FROM THE REDACTION REQUIREMENT. The redaction requirement does not apply to the following:

(1) a financial-account number that identifies the property allegedly subject to forfeiture in a forfeiture proceeding;

(2) the record of an administrative or agency proceeding;

(3) the official record of a state-court proceeding;

(4) the record of a court or tribunal, if that record was not subject to the redaction requirement when originally filed;

(5) a filing covered by Rule 5.2(c) or (d); and

(6) a pro se filing in an action brought under 28 U.S.C. §§2241, 2254, or 2255.

(c) LIMITATIONS ON REMOTE ACCESS TO ELECTRONIC FILES; SOCIAL-SECURITY APPEALS AND IMMIGRATION CASES. Unless the court orders otherwise, in an action for benefits under the Social Security Act, and in an action or proceeding relating to an order of removal, to relief from removal, or to immigration benefits or detention, access to an electronic file is authorized as follows:

(1) the parties and their attorneys may have remote electronic access to any part of the case file, including the administrative record;

(2) any other person may have electronic access to the full record at the courthouse, but may have remote electronic access only to:

(A) the docket maintained by the court; and

(B) an opinion, order, judgment, or other disposition of the court, but not any other part of the case file or the administrative record.

(d) FILINGS MADE UNDER SEAL. The court may order that a filing be made under seal without redaction. The court may later unseal the filing or order the person who made the filing to file a redacted version for the public record.

(e) PROTECTIVE ORDERS. For good cause, the court may by order in a case:

(1) require redaction of additional information; or

(2) limit or prohibit a nonparty's remote electronic access to a document filed with the court.

(f) OPTION FOR ADDITIONAL UNREDACTED FILING UNDER SEAL. A person making a redacted filing may also file an unredacted copy under seal. The court must retain the unredacted copy as part of the record.

(g) OPTION FOR FILING A REFERENCE LIST. A filing that contains redacted information may be filed together with a reference list that identifies each item of redacted information and specifies an appropriate identifier that uniquely corresponds to each item listed. The list must be filed under seal and may be amended as of right. Any reference in the case to a listed identifier will be construed to refer to the corresponding item of information.

(h) WAIVER OF PROTECTION OF IDENTIFIERS. A person waives the protection of Rule 5.2(a) as to the person's own information by filing it without redaction and not under seal.

Federal Rule of Criminal Procedure Rule 49.1—Privacy Protection for Filings Made with the Court

(a) REDACTED FILINGS. Unless the court orders otherwise, in an electronic or paper filing with the court that contains an individual's social-security number, taxpayer-identification number, or birth date, the name of an individual known to be a minor, a financial-account number, or the home address of an individual, a party or nonparty making the filing may include only:

(1) the last four digits of the social-security number and taxpayer-identification number;

(2) the year of the individual's birth;

(3) the minor's initials;

(4) the last four digits of the financial-account number; and

(5) the city and state of the home address.

(b) EXEMPTIONS FROM THE REDACTION REQUIREMENT. The redaction requirement does not apply to the following:

(1) a financial-account number or real property address that identifies the property allegedly subject to forfeiture in a forfeiture proceeding;

(2) the record of an administrative or agency proceeding;

(3) the official record of a state-court proceeding;

(4) the record of a court or tribunal, if that record was not subject to the redaction requirement when originally filed;

(5) a filing covered by Rule 49.1(d);

(6) a pro se filing in an action brought under 28 U.S.C. §§2241, 2254, or 2255;

(7) a court filing that is related to a criminal matter or investigation and that is prepared before the filing of a criminal charge or is not filed as part of any docketed criminal case;

(8) an arrest or search warrant; and

(9) a charging document and an affidavit filed in support of any charging document.

(c) IMMIGRATION CASES. A filing in an action brought under 28 U.S.C. §2241 that relates to the petitioner's immigration rights is governed by Federal Rule of Civil Procedure 5.2.

(d) FILINGS MADE UNDER SEAL. The court may order that a filing be made under seal without redaction. The court may later unseal the filing or order the person who made the filing to file a redacted version for the public record.

(e) PROTECTIVE ORDERS. For good cause, the court may by order in a case:

(1) require redaction of additional information; or

(2) limit or prohibit a nonparty's remote electronic access to a document filed with the court.

(f) OPTION FOR ADDITIONAL UNREDACTED FILING UNDER SEAL. A person making a redacted filing may also file an unredacted copy under seal. The court must retain the unredacted copy as part of the record.

(g) OPTION FOR FILING A REFERENCE LIST. A filing that contains redacted information may be filed together with a reference list that identifies each item of redacted information and specifies an appropriate identifier that uniquely corresponds to each item listed. The list must be filed under seal and may be amended as of right. Any reference in the case to a listed identifier will be construed to refer to the corresponding item of information.

(h) WAIVER OF PROTECTION OF IDENTIFIERS. A person waives the protection of Rule 49.1(a) as to the person's own information by filing it without redaction and not under seal.

Federal Rule of Bankruptcy Procedure Rule 9037—Privacy Protection for Filings Made with the Court

(a) REDACTED FILINGS. Unless the court orders otherwise, in an electronic or paper filing made with the court that contains an individual's social-security number, taxpayer-identification number, or birth date, the name of an individual, other than the debtor, known to be and identified as a minor, or a financial-account number, a party or nonparty making the filing may include only:

(1) the last four digits of the social-security number and taxpayer-identification number;

(2) the year of the individual's birth;

(3) the minor's initials; and

(4) the last four digits of the financial-account number.

(b) EXEMPTIONS FROM THE REDACTION REQUIREMENT. The redaction requirement does not apply to the following:

(1) a financial-account number that identifies the property allegedly subject to forfeiture in a forfeiture proceeding;

(2) the record of an administrative or agency proceeding unless filed with a proof of claim;

(3) the official record of a state-court proceeding;

(4) the record of a court or tribunal, if that record was not subject to the redaction requirement when originally filed;

(5) a filing covered by subdivision (c) of this rule; and

(6) a filing that is subject to §110 of the Code.

(c) FILINGS MADE UNDER SEAL. The court may order that a filing be made under seal without redaction. The court may later unseal the filing or order the entity that made the filing to file a redacted version for the public record.

(d) PROTECTIVE ORDERS. For cause, the court may by order in a case under the Code:

(1) require redaction of additional information; or

(2) limit or prohibit a nonparty's remote electronic access to a document filed with the court.

(e) OPTION FOR ADDITIONAL UNREDACTED FILING UNDER SEAL. An entity making a redacted filing may also file an unredacted copy under seal. The court must retain the unredacted copy as part of the record.

(f) OPTION FOR FILING A REFERENCE LIST. A filing that contains redacted information may be filed together with a reference list that identifies each item of redacted information and

specifies an appropriate identifier that uniquely corresponds to each item listed. The list must be filed under seal and may be amended as of right. Any reference in the case to a listed identifier will be construed to refer to the corresponding item of information.

(g) WAIVER OF PROTECTION OF IDENTIFIERS. An entity waives the protection of subdivision (a) as to the entity's own information by filing it without redaction and not under seal.

(h) MOTION TO REDACT A PREVIOUSLY FILED DOCUMENT

(1) *Content of the Motion; Service*. Unless the court orders otherwise, if an entity seeks to redact from a previously filed document information that is protected under subdivision (a), the entity must:

(A) file a motion to redact identifying the proposed redactions;

(B) attach to the motion the proposed redacted document;

(C) include in the motion the docket or proof-of-claim number of the previously filed document; and

(D) serve the motion and attachment on the debtor, debtor's attorney, trustee (if any), United States trustee, filer of the unredacted document, and any individual whose personal identifying information is to be redacted.

(2) *Restricting Public Access to the Unredacted Document; Docketing the Redacted Document*. The court must promptly restrict public access to the motion and the unredacted document pending its ruling on the motion. If the court grants it, the court must docket the redacted document. The restrictions on public access to the motion and unredacted document remain in effect until a further court order. If the court denies it, the restrictions must be lifted, unless the court orders otherwise.

Federal Rule of Appellate Procedure Rule 25(a)(5)—Filing and Service

(a) FILING.

(5) *Privacy Protection.* An appeal in a case whose privacy protection was governed by Federal Rule of Bankruptcy Procedure 9037, Federal Rule of Civil Procedure 5.2, or Federal Rule of Criminal Procedure 49.1 is governed by the same rule on appeal. In all other proceedings, privacy protection is governed by Federal Rule of Civil Procedure 5.2, except that Federal Rule of Criminal Procedure 49.1 governs when an extraordinary writ is sought in a criminal case. The provisions on remote electronic access in Federal Rule of Civil Procedure 5.2(c)(1) and (2) apply in a petition for review of a benefits decision of the Railroad Retirement Board under the Railroad Retirement Act.

Appendix B: Methodology

Sample

This study is based on an analysis of all documents filed in the federal district, bankruptcy, and appeals courts on 37 randomly selected days in calendar year 2022.[12] Because there is not a comprehensive list of all documents filed in all courts, we could not randomly select documents directly. Instead, we randomly selected a subset of dates in 2022 and analyzed all documents filed on those dates. We set the number of dates to 37, or about 10% of the total number of days in 2022.

Approximately 97% of district and bankruptcy court documents and 99% of appellate briefs are filed on non-holiday weekdays.[13] In an effort to mirror that distribution, we randomly selected 36 dates from a list of all non-holiday weekdays and one date from a list of all weekends and federal holidays. Document filings furthermore tend to be evenly distributed across quarters.[14] Correspondingly, we randomly selected nine weekday dates from each quarter.

Using these procedures, we randomly selected the following dates in calendar year 2022:

Q1	Q2	Q3	Q4
January 18	April 2*	July 18	October 18
January 25	April 15	July 25	October 25
February 4	April 22	August 4	November 4
February 8	May 4	August 8	November 8
February 11	May 6	August 11	November 14
March 14	May 11	September 9	December 14
March 15	June 9	September 12	December 15
March 21	June 10	September 16	December 21
March 30	June 16	September 27	December 27
	June 28		

*Weekend day

Dataset

To construct our dataset, we first downloaded PDFs of the 4,681,055 documents filed in the federal district, bankruptcy, and appeals courts on the 37 dates in our sample. For the purposes of this study, we considered a document to be the entire contents of a single PDF filed with the court.[15] We then used the Python library PyPDF to convert the PDFs into

[12] In contrast, the 2010 and 2015 Center studies were based on nonprobability samples. The 2010 study examined all documents filed in district and bankruptcy courts in November and December of 2009. The 2015 study examined all documents filed in district and bankruptcy courts in November 2013.
[13] Tim Reagan, et al., "Electronic Filing Times in Federal Courts," Federal Judicial Center, April 25, 2022, https://www.fjc.gov/content/365889/electronic-filing-times-federal-courts.
[14] Ibid.
[15] Some PACER docket entries contain multiple filings, with each being an individual downloadable PDF.

searchable text files. PDFs that could not be converted using PyPDF were converted using the Tesseract OCR engine in Python. Of the 4,681,055 PDFs we downloaded, 4,674,242 (99.9%) were successfully converted into searchable text files. The vast majority (95%, 6,456) of PDFs that could not be converted were documents from appellate cases.

Next, we ran a Python script that extracted nine-digit numbers from the text files, along with the 200 characters that preceded and followed the numbers. We also extracted information about each document and case, including the court name, division, docket number, docket entry, and docket sequence numbers. We used this information to create 292 spreadsheets: one for each of the 94 district courts; one for each of the 89 unconsolidated bankruptcy courts, as well as individual spreadsheets for bankruptcy filings in the Eastern and Western Districts of Arkansas (which share a bankruptcy court but docket cases separately) and for the three territorial courts;[16] one for each of the 12 regional courts of appeals; and one for each of the 89 unconsolidated bankruptcy courts with proof of claim registers, as well as one each for the proof of claim registers in the Eastern and Western Districts of Arkansas and the territorial court in Guam.[17]

Each row of these spreadsheets represented either an instance of a nine-digit number found in the documents or a single entry for a document in which no nine-digit numbers had been found. The full dataset contained 30.2 million rows. We discovered that about 21.6 million of these rows were related to a particular type of nine-digit number that appeared regularly in 3M Products Liability Litigation (MDL No. 2885) cases filed in the Northern District of Florida. This number was not a valid SSN, so these rows were omitted. We also found that 4.2 million rows represented documents with no identified nine-digit numbers. The remaining 4.4 million rows included nine-digit numbers that we analyzed further to determine if they were valid SSNs.

Search Algorithm Development and Validation

We developed a search algorithm in the R programming language to help us identify which of the 4.4 million nine-digit numbers were mostly likely to be valid SSNs.

To begin, a team of researchers manually inspected documents that contained 123,911 identified numbers (rows) across 27 district court datasets and labeled them as valid or invalid SSNs. We observed that valid SSNs tended to appear in predictable contexts or formats. We used these patterns to write an algorithm that predicted whether a row was likely a tax identification number (TIN), possibly a TIN, or likely not a valid TIN.

The algorithm predicted that a nine-digit number was “likely” or “possibly” a TIN if any of the following conditions were met:

[16] Bankruptcy cases in the district courts of Guam, the Virgin Islands, and the Northern Mariana Islands are heard by district court judges or visiting bankruptcy judges.
[17] The territorial courts of the Virgin Islands and the Northern Mariana Islands did not have any proof of claim filings on the dates in the sample.

- **Number appeared in a common TIN context.** A row was labeled LIKELY TIN if the number appeared within eight characters of any of the following strings (not case sensitive):

"EIN," "Employer Identification," "Employer Identification No," "Employer ID," "Employer I.D," "Employer 1D," "Employer 1.D," "Employer Identification Number," "Employer Number," "Employer ID Number," "Employee Identification Number," "Tax ID," "Tax I.D," "tax identification number," "tax identification," "tax identification no," "Tax ID#," "Tax#," "Tax ID Number," "Tax I.D. Number," "Tx ID," "Tx I.D," "TaxID," "Tax. ID," "Tax1D," "Tax 1D," "Tax 1.D," "Taxpayer ID," "Taxpayer I.D," "Taxpayer ID No," "Taxpayer ID Number," "Taxpayer I.D. Number," "Taxpayer ID#," "Taxpayer 1D," "Taxpayer 1.D," "Taxpayer Number," "Taxpayer No," "Taxpayer Identification," "Taxpayer Identification Number," "Taxpayer Identification Number (US)," "IRS," "IRS No," "IRS Number," "Internal Revenue Service," "Internal Revenue Service Number," "I.R.S," "I.R.S. Number," "I.R.S. No," "FEIN," "ITIN," "EID," "TID," "ATIN," "PTIN," "TIN," "FIN," "SSI," "S.S.I," "SSI Number," "SSI No," "S.S.I. Number," "SSI ID," "SS Number," "SS No," "S.S. No," "S.S. NUMBER," "SS#," "SS Nbr," "SSA," "SSA Number," "Social Security," "Social Security No," "Social Security Number," "social security account number," "social security acct no," "social security account no," "SSN," "SSN/SIN," "*SSN," "(SSN)," "[SSN," "SS," "'SS," "(SSN," "8.8.N," "soc. sec. no," "SOC.SEC," "soc sec," "soc. sec," "socsec," "SOC."

- **Number appeared in a common TIN format.** A row was labeled LIKELY TIN if it followed either of these formats: 123-45-6789 and 12-3456789.

- **Number appeared in a less common TIN format.** A row was labeled POSSIBLE TIN if it followed either of these formats: 123.45.6789 and 123 45 6789.

- **The same number matched a previous condition.** In the last step, the algorithm copied the number strings and then removed all punctuation and spaces from the strings so they appeared in the same format. For example, the numbers 123-45-6789, 123 45 6789, and 123456789 were all formatted to appear as 123456789. The algorithm then sorted and grouped the resulting standardized numbers. If any member of a group had previously been labeled LIKELY TIN or POSSIBLE TIN, all other members of the group were also labeled as such. For example, if the number 123456789 appeared in four rows and it was labeled LIKELY TIN in one row because it had appeared after the term "SSN#," the other three rows would be updated to reflect that they were also LIKELY TIN.

Finally, we ran multiple tests to validate the algorithm's predictions. Human coders who were assisted by the algorithm's predictions identified an estimated 99% of valid SSNs in the district court data, 99% in the bankruptcy court data, and 100% in the appeals court data. By comparison, human coders working without the assistance of the algorithm's

predictions found 92% of valid SSNs in the district court data, 97% in the bankruptcy court data, and 83% in the appeals court data. The search algorithm therefore not only made the process of identifying SSNs more efficient, it also improved accuracy.

Manual Coding of SSNs

The search algorithm predicted that 51,894 of the 4.4 million nine-digit numbers could be valid tax identification numbers. To make a final determination, each of those observations that had been flagged by the algorithm were double-coded by researchers who independently inspected each row. In many cases, researchers referenced the original document to view the number in context. Researchers coded observations as "SSN," "ITIN," "EIN," "TIN Unspecified," or "Not Valid." Researchers also had the option of using the code "Follow Up" for any observations they were unsure about. In most cases, the two coders assigned the same label. When the coders disagreed or when one or both coders labeled an observation "Follow Up," senior members of the research team attempted to make a final determination to the extent possible. This process identified 22,391 SSNs and ITINs.

Manual Coding of Exemptions

Next, for each case with an identified SSN, data from the Center's Integrated Database (IDB)[18] were linked and used to flag possible exemptions and waivers. Cases were flagged as potentially exempt if they were removals from state court, social security cases, civil immigration cases, habeas corpus cases with a pro se party, or administrative agency cases or appeals. Cases were flagged as potential waivers if they included one or more pro se parties.

All 22,391 SSNs and ITINs were then double-coded by researchers who independently inspected each row to determine whether the number was or was not exempt under the Privacy Rules. Some numbers were exempt for multiple reasons. We noted each of these reasons using the exemption codes below. Disagreements between coders were inspected and resolved by a senior member of the research team.

We interpreted the exemption provisions of the privacy rules broadly and generally counted unredacted SSNs as exempt if a filing party could have reasonably understood the rules as providing an exemption. We used an expansive understanding of the terms "official record" and "state-court proceedings" to include any document that appears to be all or part of a record of any type of proceeding from a state court. We also interpreted the criminal rules as exempting SSNs appearing in non-federal charging documents filed in criminal proceedings in federal court. Finally, we treated SSNs found in attachments to warrants and charging documents as exempt under the criminal rules.

[18] The IDB contains data on civil case and criminal defendant filings and terminations in district, bankruptcy, and appellate courts and associated case information from 1970 to the present. The Center receives regular updates of the case-related data as routinely reported by the courts to the AO. The Center then post-processes the data, consistent with the policies of the Judicial Conference governing access to these data, into a unified longitudinal database, the IDB. It is available here: https://www.fjc.gov/research/idb

Exemption Codes

Miscellaneous
1 = Record of a state court proceeding
14 = Documents filed under seal

Pro se documents
2 = Filer included own SSN (suggesting waiver of the privacy protections)

Criminal documents (including attachments)
5 = Criminal charging document/affidavit
6 = Criminal arrest/search warrant
7 = Criminal investigation or other document prepared prior to filing of criminal charge

Bankruptcy documents
8 = Non-attorney bankruptcy petition preparer (e.g., Bankruptcy Form 119)

Appeals documents
9 = Filing in appeal of Railroad Retirement Board benefits decision

Civil documents
4 = Pro se party filing in a habeas corpus proceeding under 28 U.S.C. §§ 2241, 2254, or 2255
10 = Filing in civil social security case (i.e., action for benefits under the Social Security Act)
11 = Record of an administrative agency proceeding (except in bankruptcy cases if record filed with proof of claim)
12 = Immigration case (i.e., action relating to immigration removal, relief from removal, benefits, or detention)
13 = Record of a court or tribunal, if that record was not subject to the redaction requirement when originally filed