**2024 Report of the Judicial Conference of the United States
on the Adequacy of Privacy Rules Prescribed
Under the E-Government Act of 2002**

**PREPARED FOR THE
U.S. SENATE AND HOUSE OF REPRESENTATIVES**

**JUDICIAL CONFERENCE OF THE UNITED STATES**

**June 2024**

2024 REPORT OF THE JUDICIAL CONFERENCE OF THE UNITED STATES
ON THE ADEQUACY OF PRIVACY RULES PRESCRIBED
UNDER THE E-GOVERNMENT ACT OF 2002

The E-Government Act of 2002 directed the judiciary to promulgate rules, under the Rules Enabling Act, "to protect privacy and security concerns relating to electronic filing of documents and the public availability … of documents filed electronically."  Pub. L. No. 107-347, 116 Stat. 2914, § 205(c)(3)(A)(i).  The privacy rules – Appellate Rule 25(a)(5), Bankruptcy Rule 9037, Civil Rule 5.2, and Criminal Rule 49.1 – took effect on December 1, 2007.

Subject to specified exemptions, the privacy rules require that filers redact from documents filed with the court (1) all but the last four digits of an individual's social-security number or taxpayer-identification number (these numbers are collectively referred to here as the SSN); (2) the month and day of an individual's birth; (3) all but the initial letters of a known minor's name; (4) all but the last four digits of a financial-account number; and (5) in criminal cases, all but the city and state of an individual's home address.  In recognition of the pervasive presence of sensitive personal information in filings in actions for benefits under the Social Security Act, and in proceedings relating to an order of removal, to relief from removal, or to immigration benefits or detention, the privacy rules exempt filings in those matters from the redaction requirement but also limit remote electronic access to those filings.

Section 205(c)(3)(C) of the E-Government Act directs that, every two years, "the Judicial Conference shall submit to Congress a report on the adequacy of [the privacy rules] to protect privacy and security." This report covers the period from June 2022 to June 2024.

The report proceeds in four parts.  Part I discusses potential rule amendments (i) under consideration by the rules committees at the time of the 2022 Report, or (ii) added to the rules committee dockets since the 2022 Report was completed.   Part II discusses ongoing implementation efforts by the Administrative Office of the United States Courts (the AO), the Federal Judicial Center (the FJC), and others to protect privacy in court filings and opinions.  Part III discusses research undertaken by the FJC to assess adherence to the privacy rules.  Part IV concludes with a summary and an overview of anticipated next steps.

I.      **Potential Privacy-Related Rules Amendments Under Consideration by the Rules Committees Since June 2022.**

This section addresses topics under consideration by the rules committees at the time of the 2022 Report or added to the committees' agendas since that report was completed.  Part I.A. discusses potential amendments to Criminal Rule 49.1.  Part I.B. discusses ongoing deliberations concerning applications to proceed in forma pauperis, or without prepayment of fees, in appeals.  Part I.C. notes proposals to adopt a Civil Rule addressing the sealing of court filings.  Part I.D. discusses proposals to require the full redaction of SSNs in court filings and to restrict the dissemination of an individual's full SSN to creditors in bankruptcy cases, and Part I.E. discusses two new suggestions proposing changes to the civil rules to address privacy and cybersecurity risks in civil litigation.

A.	**Potential Amendments to Criminal Rule 49.1**

At the time of the 2022 Report, the Criminal Rules Committee was evaluating whether any change to Criminal Rule 49.1 is needed to address a reference – in the 2007 committee note to that Rule – to the March 2004 "Guidance for Implementation of the Judicial Conference Policy on Privacy and Public Access to Electronic Criminal Case Files" from the Committee on Court Administration and Case Management (CACM). The Committee's consideration of a change was prompted by a public suggestion questioning whether the guidance, as outlined in the note, is consistent with caselaw concerning rights of public access to information contained in criminal defendants' CJA applications. Since the 2022 Report was issued, the Committee concluded that an amendment to Criminal Rule 49.1 would not change the note's reference to the CACM Committee's March 2004 guidance and that an amendment is otherwise not warranted.

In March 2024, the U.S. Department of Justice submitted a suggestion to the Criminal Rules Committee proposing an amendment to Rule 49.1 to require that all publicly available court filings refer to minors by pseudonyms rather than by their initials. The Committee's work on this matter is at an early stage. A new Rule 49.1 Subcommittee has been formed to study this proposal. If the Criminal Rules Committee concludes that an amendment to Criminal Rule 49.1 is warranted, the other advisory committees would then consider whether parallel amendments to the other privacy rules would be appropriate.

B.	**Potential Amendments Concerning Applications to Proceed In Forma Pauperis (IFP)**

The Appellate Rules Committee has been considering suggestions to revise Appellate Form 4 (Affidavit Accompanying Motion for Permission to Appeal In Forma Pauperis). The basic suggestion is that Form 4 could be substantially simplified while still providing the courts of appeals with enough detail to decide whether to grant IFP status. At its April 2024 meeting, the Appellate Rules Committee recommended for publication and public comment proposed amendments to Form 4 that would reduce the amount of personal financial detail the form requires. If publication goes forward as recommended, and the proposed amendments receive subsequent approvals in the ordinary course, a revised version of the form could go into effect as early as December 1, 2026.

C.	**Proposals to Adopt a Rule on Sealing of Court Filings**

The Civil Rules Committee has before it proposals to adopt a rule setting standards and procedures governing the sealing of court filings. The Committee has referred these proposals to its Discovery Subcommittee for initial evaluation. The subcommittee has recently started an information-gathering effort to identify logistical issues that might arise if some of the proposed measures in the suggestions for sealing standards were to be adopted.

D.	**Proposals for Further Restrictions on the Use of SSNs**

Since the 2022 Report, the rules committees have received a suggestion to require full redaction of SSNs in court filings, and the Bankruptcy Rules Committee has received suggestions

to eliminate the debtor's partially redacted SSN and address information on some of the notices filed on the court docket and to stop sending the debtor's full SSN to creditors in a bankruptcy case.

### D.1    Suggestion from Senator Ron Wyden

As noted in the 2022 Report, in 2015-2016, the Appellate, Bankruptcy, Civil, and Criminal, Rules Committees considered suggested amendments to the privacy rules that would require redaction of an individual's entire SSN in court filings.  In evaluating the proposal, participants noted that the rules committees had considered full redaction of such numbers when formulating the privacy rules, but had concluded that the last four digits were needed in bankruptcy proceedings to confirm debtor identity.  Given the E-Government Act's requirement to promulgate rules that are uniform "to the extent practicable" in protecting privacy and security issues,[1] the Appellate, Civil, and Criminal Rules Committees followed the lead of the Bankruptcy Rules Committee in requiring redaction of all but the last four digits of an individual's SSN. Based on continued agreement with that analysis, the advisory committees decided not to propose amendments to the privacy rules at that time.

In an August 4, 2022, letter concerning a draft of the 2022 Report, Senator Ron Wyden suggested that the rules committees reconsider a proposal to redact the entire SSN from court filings.  The Bankruptcy Rules Committee took the lead in considering Senator Wyden's suggestion at its spring 2023 meeting.

By way of background, in the 1990s, the judiciary considered privacy concerns related to the increasing ease of access to electronic public records through the internet.  The CACM Committee – with input from other Judicial Conference Committees, particularly the Bankruptcy Rules Committee, as well as the public – recommended a privacy policy governing the electronic availability of case file information, which reflected a careful balance between public access and individual privacy.  The Judicial Conference adopted this policy in 2001 (JCUS-SEP/OCT 2001, pp. 48-50).  Among other things, the policy required the modification or partial redaction of SSNs in civil case files and directed the Bankruptcy Rules Committee to amend the rules as necessary to allow a court to collect a debtor's full SSN but display only the last four digits.  Under this policy, several amendments to the bankruptcy rules and forms were implemented in 2003 to limit disclosure of a party's SSN or other personally identifiable information.  The bankruptcy petition forms, and Official Form 416A, Caption (Full), were modified to include only the last four digits of a debtor's SSN in order "to afford greater privacy to the individual debtor, whose bankruptcy case records may be available on the Internet." *See* 2003 committee notes to Official Bankruptcy Forms 101, 105, and 416A.  Rule 1005 was similarly amended to require only the last four digits of the debtor's SSN in the caption of a petition.  At the same time, Rule 2002(a)(1) was amended to require that the debtor's full SSN be included in the official form providing notice of the bankruptcy case that is sent to creditors under 11 U.S.C. § 341 or § 1104(b), but that the filed version of the form include only the last four digits of the SSN.  As explained in the committee note (2003) to Rule 2002:

---

[1] E-Government Act of 2002, Pub. L. No. 107-347, § 205(c)(3)(A)(ii).

This will enable creditors and other parties in interest who are in possession of the debtor's social security number to verify the debtor's identity and proceed accordingly. The filed Official Form 9, however, will not include the debtor's full social security number. This will prevent the full social security number from becoming a part of the court's file in the case, and the number will not be included in the court's electronic records. Creditors who already have the debtor's social security number will be able to verify the existence of a case under the debtor's social security number, but any person searching the electronic case files without the number will not be able to acquire the debtor's social security number.

All versions of Official Form 9 (now Official Forms 309A-309I) were amended accordingly to include only the last four digits of the debtor's SSN in the official copy included in the case file.

The Bankruptcy Rules Committee's spring 2023 minutes reflect that in considering Senator Wyden's suggestion, members noted that two statutory provisions preclude a rule change that would require the full redaction of SSNs in all filings. Section 110(c) of the Bankruptcy Code requires bankruptcy petition preparers to include their full SSN on any bankruptcy filing they have prepared for filing in the case. And § 342(c) requires that the last four digits of the debtor's SSN be included on notices "required to be given by the debtor to a creditor under this title, any rule, any applicable law, or any order of the court." Outside those statutory constraints, however, the Committee is considering related suggestions that would remove the debtor's partially redacted SSN on some notices sent under Rule 2002, and it is evaluating the need for the partially redacted SSN on some bankruptcy forms where it is currently required. Those proposals are discussed in Part D.2 below.

A working group composed of the rules committees' reporters is also in the beginning stages of considering whether, despite the E-Government Act preference for uniform privacy rules, the rules committees should reconsider fully redacting SSNs from filings in civil and criminal cases irrespective of the need for full or partially redacted SSNs in some bankruptcy filings. (The appellate privacy rule incorporates the privacy rule of the type of case – bankruptcy, civil, or criminal – that is being appealed.) At the spring 2024 meetings of the advisory committees, the working group provided a sketch for a possible amendment to require the full redaction of SSNs in court filings but recommended that such an amendment to the Civil and Criminal Rules should not be taken up in isolation but should be part of a more comprehensive review of the privacy rules. The working group will continue to work with the advisory committees to identify areas of common concern and to assist in coordination of proposed changes.

### D.2 *Suggestions That Would Remove Redacted SSNs From Some Bankruptcy Notices and Forms.*

Bankruptcy Rule 1005 requires that the caption of the petition contain the name of the court, title of the case, and docket number. It further requires that the title of the case include the debtor's name, employer identification number, last four digits of the debtor's SSN, and all other

names used by the debtor within eight years before filing the petition. Bankruptcy Rule 2002(n) requires that the caption of every notice given under Rule 2002 comply with Rule 1005.

In 2023, the Bankruptcy Rules Committee received a suggestion from a group of bankruptcy clerks from the Eighth Circuit suggesting that Rule 2002(n) be amended to eliminate the requirement that the caption of every notice given under Rule 2002 comply with Rule 1005. The AO's Bankruptcy Clerks Advisory Group submitted a second suggestion supporting the clerks' suggestion.

The bankruptcy clerks state that the caption requirements "are substantial and can add a significant amount of length, and therefore cost, to a Rule 2002 notice." They also note that, despite the requirements of Rule 2002(n), there is a long-standing practice of bankruptcy clerks in their circuit to provide the Rule 1005 caption requirements only on the Notice of Bankruptcy Case. Thereafter, the clerk's office uses a shorter caption that "generally follows Official Form 416B" which requires only the debtor's name, and the bankruptcy case and chapter numbers. If the suggestion is adopted, most notices under Rule 2002 would no longer include a field for the debtor's partially redacted SSN. A subcommittee of the Bankruptcy Rules Committee, with the help of the FJC, has surveyed bankruptcy clerks about the desirability of including all the information required by Rule 1005 in routine notices under Rule 2002.

In addition, in connection with Senator Wyden's suggestion, the subcommittee, with the help of the FJC, has begun to survey debtor attorneys, chapter 7, 12, and 13 trustees, creditor attorneys, various tax authorities and representatives of the National Association of Attorneys General about whether bankruptcy forms that currently require inclusion of the debtor's redacted SSN must or should continue to do so.

### D.3    Suggestion 23-BK-A to Restrict Dissemination of the Debtor's Full SSN

A staff attorney for a chapter 13 trustee, suggested that Bankruptcy Rule 2002(a)(1) be amended to stop sending the debtor's full SSN to creditors. Similar suggestions were received in 2011 and 2015. In considering the earlier suggestions, although Committee members recognized the importance of protecting debtors from improper disclosure of their full SSN, they also recognized that creditors such as the IRS rely on the full SSN to ensure that they are seeking payment from the correct debtor or to determine whether a debtor from whom they are seeking payment has filed for bankruptcy protection. A subcommittee reviewing the suggestion noted that some creditors continue to use the full SSN to ensure accurate debtor identification. The subcommittee therefore recommended no changes. The Bankruptcy Rules Committee discussed the recommendation at its spring 2023 meeting and decided to take no action on the suggestion.

### E.    Proposals to Amend the Civil Rules to Further Protect Privacy Rights and Prevent Cybersecurity Problems

In September 2023, the Lawyers for Civil Justice (LCJ) submitted a suggestion for the comprehensive examination of the Civil Rules and to implement a framework for the court and parties to protect privacy rights and prevent cybersecurity problems at various stages of civil litigation, including discovery. LCJ identified a number of Civil Rules for potential amendments

to better protect parties and non-parties from disclosure of personal and confidential information. In November 2023, a private attorney wrote to the rules committees in support of LCJ's proposal. His submission encouraged the Civil Rules Committee to address comprehensively the privacy and cybersecurity risks in civil litigation. The Committee is in the early stages of considering these suggestions.

## II.    Ongoing Implementation Efforts to Protect Privacy in Court Filings and Opinions

As mentioned above, the privacy rules require that the filer redact certain personal identifiers from court filings. Additionally, due to the pervasive presence of sensitive personal information in Social Security and immigration cases, the privacy rules exempt filings in those matters from the redaction requirement but also limit remote electronic access to those filings. The opinions in these cases, however, are widely available to the public via PACER and other legal research databases that are easily searchable. The CACM Committee and the AO have recently engaged in a number of outreach and educational efforts to protect personal information.

In May 2023, the CACM Committee sent a memorandum to the courts sharing suggested practices to protect personal information in court filings and opinions. With regard to court filings, the memorandum urged the courts to continue or to consider initiating outreach efforts to litigants and members of the bar to ensure they are aware of redaction obligations and the need to minimize the appearance of private identifiers in certain court filings.[2]

The May 2023 memorandum also reminded the courts about a possible concern regarding sensitive personal information in Social Security and immigration opinions and a suggested practice of using only the first name and last initial of any non-government parties in the opinions.[3] Since this suggested practice was first shared with the courts in 2018, many courts have redacted party names in their opinions. In addition, some districts have adopted a local rule or internal operating procedure addressing the practice. Finally, the May 2023 memorandum reminded courts about a software change implemented by the AO in 2020 that masks information such as case and party names in extracts of Social Security and immigration opinions provided to the Government Printing Office and the GovInfo database for publication.

Beyond sharing suggested practices directly with the courts, the CACM Committee recently requested that the AO and FJC explore other ways to increase awareness about ways to protect privacy in court filings and opinions. The AO recently updated several sections of the

---

[2] Specifically, similar to a memorandum sent to courts by the CACM Committee in November 2011, the memorandum emphasized that courts should ensure they are aware of (1) filers' redaction obligations under the privacy rules; (2) measures to minimize the appearance of private identifiers in court filings; (3) the obligation to secure a court order before redacting information beyond that specifically identified in the privacy rules; and (4) the obligation to redact private identifiers from transcripts of proceedings.

[3] This suggested practice was developed following extensive consultation with stakeholders inside and outside the judiciary as a way to balance the need to provide public access to Social Security and immigration opinions while protecting personal information. The CACM Committee first shared this suggested practice in a May 2018 memorandum to the courts.

judiciary's internal and public websites to include updated information regarding privacy rule requirements and suggested practices. Furthermore, the FJC is exploring ways to increase references to these suggested practices in its educational materials and trainings for new judges, court unit executives, and law clerks, and it will explore developing a model webpage that courts can include on their local websites to increase awareness among the bar and the public.

Additionally, the current case management system continues to notify filers via a prominent banner titled "Redaction Agreement" that appears immediately after a filer logs in to remind them of the redaction requirements in the Appellate, Bankruptcy, Civil, and Criminal Rules, and that the requirements apply to all documents, including attachments. To proceed, the filer is required to check a box acknowledging that they have read the notice and understand their obligation to comply with the redaction requirements. Thereafter, before a filer electronically submits a document to the court, the system presents a reminder asking "have you redacted?"

Finally, the CACM Committee has urged the AO to implement features in the modernized case management system to automate and facilitate a litigant's review of court filings for compliance with the redaction requirements in the Appellate, Bankruptcy, Civil, and Criminal Rules. The CACM Committee will continue to explore other possible ways to protect private information in court filings and opinions.

## III. Federal Judicial Center Research on Unredacted Personal Information

As noted in prior reports on the adequacy of the privacy rules, the FJC has undertaken several studies of compliance with the redaction requirements. The FJC in 2010 conducted a survey of federal court filings to ascertain how often unredacted SSNs appeared in those filings.[4] In 2015, the FJC reported the results of its follow-up study on the same topic.[5] The follow-up study searched 3,900,841 documents filed during a one-month period in late 2013 and found that 5,437 (or less than 0.14 percent of the documents) included one or more unredacted SSNs. This is a greater percentage than was found in the 2010 study; but the 2015 study explained that the difference was due to an improvement in search methodology. In the 2015 study, the researchers reprocessed the documents using optical character recognition (OCR), which enabled them to identify SSNs in documents that were originally filed in non-text-searchable format. The researchers noted that, because OCR had not been used for the 2010 study, that study had failed to reflect the full incidence of unredacted SSNs. They observed that a comparison of the two studies' findings, taking into account the difference in methodologies, "suggests that the federal courts have made progress in recent years in reducing the incidence of unredacted Social Security numbers in federal court documents, especially in bankruptcy court documents."[6]

---

[4] See Memorandum from George Cort & Joe Cecil, Research Division, FJC, to the Privacy Subcommittee of the Judicial Conference Committee on Rules of Practice and Procedure, Social Security Numbers in Federal Court Documents (April 5, 2010).

[5] See Joe S. Cecil et al., Unredacted Social Security Numbers in Federal Court PACER Documents (FJC 2015).

[6] *Id.* at 11.

In January 2023, the CACM Committee asked the FJC to update its 2015 study of court filings for adherence to the privacy rules. The FJC's updated study, completed in May 2024 and attached as Exhibit 1, used an expanded sampling procedure, more advanced methodology, and context-specific exemption coding, which limit the ability to make direct comparisons to the 2010 and 2015 studies.

For the updated study, the FJC downloaded and analyzed all documents (4,674,242) filed in the district courts (2,017,908), bankruptcy courts (2,518,202) (including proof of claim filings), and appeals courts (138,132) on 37 randomly selected days in calendar year 2022. The FJC searched these documents for possible instances of unredacted SSNs, and identified 22,391 unredacted SSNs belonging to approximately 8,300 individuals. Of the nearly 4.7 million documents analyzed, just 4,525 (0.10%) contained one or more unredacted SSNs.[7] Moreover, within the set of unredacted SSNs, the FJC concluded that approximately 22% appear to be exempt from the redaction requirement and an additional 6% belong to pro se parties who waived the privacy protections by filing their own SSN in an unsealed document. The FJC analysis also indicates that a large percentage of the unredacted SSNs occurred in a relatively small number of documents. For example, 45% of the unredacted SSNs (10,042) were found in 17 documents, with just two documents in the same case accounting for nearly 6,200 unredacted SSNs.[8]

In future studies, the FJC intends to report on instances of unredacted private information beyond social-security numbers in court filings. For instance, the FJC will identify documents with unredacted birth dates, minor names, financial account numbers, and (in criminal cases) details of an individual's home address. The FJC also intends to analyze Social Security and immigration opinions for the presence of full names of non-government parties. The FJC will collaborate with the AO to assist with future reports to Congress on the adequacy of the privacy rules.

## IV. Conclusion

In the two years since the Judicial Conference's 2022 Report to Congress on the adequacy of the privacy rules, the rules committees have considered several proposed rule changes that include privacy-related issues. As described in Part I, the Bankruptcy, Civil, and Criminal Rules Committees are reconsidering the need for the last four digits of SSNs in court filings, and they are also considering whether the privacy rules need to remain uniform with respect to the level of redactions applied to SSNs. One suggestion noted in the 2022 Report, proposed amendments to Appellate Form 4, is now on track to be published for comment in 2024, while several more recent privacy-related suggestions are in the beginning stages of consideration. Part II describes ongoing implementation efforts to protect privacy in court filings and opinions. Among other things, the CACM Committee sent a memorandum to the courts in May 2023 sharing suggested practices to protect privacy and encouraging continued outreach and educational efforts. The memorandum

---

[7] The breakdown of unredacted SSNs by court was as follows: district court: 0.12%, bankruptcy court: 0.07%, court of appeals: 0.17%.

[8] In this example, a civil case, a party filed a single document containing 3,099 SSNs twice, using a "redaction" method that is easily circumvented.

also reminded courts about the possible inclusion of sensitive information in Social Security and immigration opinions and reminded courts of a software fix implemented in 2020 that can mask certain information in extracts of Social Security and immigration opinions.  Part II also reports that the CACM Committee has asked the AO and FJC to explore other ways to increase awareness of the need to protect privacy in court filings and opinions, leading to updates in the judiciary's internal and external websites, and efforts by the FJC to address privacy issues in educational materials for new judges.  Part III, in turn, discusses the FJC's 2024 update of its studies in 2010 and 2015 concerning the prevalence of unredacted SSNs in court filings.  With respect to SSNs, the FJC's 2024 study reveals that non-compliance with the existing privacy rules remains very low.  Upcoming FJC studies addressing other aspects of the privacy rules will be considered by the rules committees and the CACM Committee in the coming years and will be addressed in future privacy reports.

# Exhibit 1

**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

*This report was produced at U.S. taxpayer expense.*

# 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**

o   Record of a state court proceeding

o   Pro se party filing in a habeas corpus proceeding under 28 U.S.C. §§ 2241, 2254, or 2255

o   Criminal charging document/affidavit

o   Criminal arrest/search warrant

o   Criminal investigation or other document prepared prior to filing of criminal charge

o   Non-attorney bankruptcy petition preparer (e.g., Bankruptcy Form 119)

o   Filing in appeal of Railroad Retirement Board benefits decision

o   Filing in civil social security case (i.e., action for benefits under the Social Security Act)

o   Record of administrative agency proceeding (except in bankruptcy cases if record filed with proof of claim)

o   Immigration case (i.e., action relating to immigration removal, relief from removal, benefits, or detention)

o   Record of a court or tribunal, if that record was not subject to the redaction requirement when originally filed

o   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