2015

Privacy and Court Records: An Empirical Study

David S. Ardia
University of North Carolina School of Law, ardia@email.unc.edu

Anne Klinefelter

Publication: *Berkeley Technology Law Journal*

PRIVACY AND COURT RECORDS: AN EMPIRICAL STUDY

David S. Ardia[†] *& Anne Klinefelter*[††]

ABSTRACT

As courts, libraries, and archives move to make court records available online, the increased ease of public access raises concerns about privacy. Little work has been done, however, to study how often sensitive information appears in court records and the context in which it appears. This Article fills this gap by analyzing a large corpus of briefs and appendices submitted to the North Carolina Supreme Court from 1984 to 2000. Based on a survey of privacy laws and privacy scholarship, we created a taxonomy of 140 types of sensitive information, grouped into thirteen categories. We then coded a stratified random sample of 504 court filings in order to determine the frequency of appearance of each sensitive information type and to identify relationships, patterns, and correlations between information types and various case and document characteristics.

We present several important findings. First, although a wide variety of sensitive information appears in the court records we sampled, it is not uniformly distributed throughout the records. Most of the documents contained relatively few incidences of sensitive information while a handful of documents contained a large number of pieces of sensitive information. Second, court records vary substantially in the types and frequency of sensitive information they contain. Sensitive information in seven of the categories—"Location," "Identity," "Criminal Proceedings," "Health," "Assets," "Financial

---

DOI: http://dx.doi.org/10.15779/Z38TR9C

† Assistant Professor of Law, University of North Carolina School of Law, and Faculty Co-Director, UNC Center for Media Law and Policy.

†† Associate Professor of Law, University of North Carolina School of Law, and Director, Kathrine R. Everett Law Library.

We could not have completed this project without the help of a dedicated team of researchers. Particular thanks is due to Esther Earbin, who oversaw the work of our coders: Dylan Arant, Nancy Brown, Alex Contarino, Kerry Dutra, Flora Feng, Emma Gilmore, Wilson Hood, Melissa Reed Muse, Haley Shell, and Josh Smith. Thank you also to Guangya (Ya) Liu for empirical research support, to Jesse Griffin and Gary Wilhelm for technical support, and to Catherine Caycedo, Daniel Parisi, Maddie Salamone, Dave Hansen, and Kate Dickson for research assistance. We are also grateful to Tamar Birckhead, John Conley, Woodrow Hartzog, Christopher Hoofnagle, William Marshall, Robert Mosteller, Helen Nissenbaum, Cathy Packer, Mary-Rose Papandrea, and the participants at the BCLT/BTLJ symposium on "Open Data: Addressing Privacy, Security, and Civil Rights Challenges" for their valuable comments and suggestions. The collection and coding of data was funded by Microsoft Corporation and the Berkeley Center for Law & Technology.

Information," and "Civil Proceedings"—appeared much more frequently than the other categories in our taxonomy. Third, information associated with criminal proceedings, such as witness and crime victim names, is pervasive in court records, appearing in all types of cases and records. Fourth, criminal cases have disproportionately more sensitive information than civil or juvenile cases, with death penalty cases far exceeding all other case types. Fifth, appendices are generally not quantitatively different from legal briefs in terms of the frequency and types of sensitive information they contain, a finding that goes against the intuition of many privacy advocates. Sixth, there were no overarching trends in the frequency of sensitive information during the seventeen-year period we studied.

Although we found a substantial amount of sensitive information in the court records we studied, we do not take a position regarding what information, if any, courts or archivists should redact or what documents should be withheld from online access or otherwise managed for privacy protection. These largely normative questions must be answered based on a careful balancing of the competing public access and privacy interests. Nevertheless, we expect that this highly granular view of the occurrence of sensitive information in these North Carolina Supreme Court records will help policymakers and judges evaluate the potential harms to privacy interests that might arise from online access to court records. We also hope that scholars will draw on our taxonomy and empirical data to develop and ground normative arguments about the proper approach for balancing government transparency and personal privacy.

TABLE OF CONTENTS

I. INTRODUCTION

Courts across the country are moving quickly to digitize their records and make them available online.[1] Some courts are doing this work themselves, while others are relying on third parties, such as libraries and archives, to make public access possible. All, however, are dealing with one central and unavoidable issue: privacy.[2]

Court records contain a variety of types of information that could be characterized as "private" or "sensitive,"[3] ranging from social security numbers to the names of minor children involved in sexual abuse. In *State v. Bright*, for example, a brief filed by the State of North Carolina describes the abduction and rape of a ten-year-old girl, naming the child in full on the first page and continuing to identify her by first name on nearly every subsequent page of the brief.[4] Similarly, in *Dean v. Cone Mills Corporation*, the plaintiff-appellant's petition for discretionary review to the North Carolina Supreme Court includes an appendix comprising the plaintiff's voluminous medical file and contains multiple references to his social security number, date of birth, and home address.[5]

Little work has been done, however, to study how often sensitive information appears in court records and the context in which it appears.[6]

---

1. *See* NATIONAL CENTER FOR STATE COURTS, TRENDS IN STATE COURTS (2014), http://www.ncsc.org/~/media/Microsites/Files/Future%20Trends%202014/2014%20NCSC%20Trends%20Report.ashx (highlighting state courts' efforts to move to e-filing and the conversion of paper case documents into digital images); Paul H. Anderson, *Future Trends in Public Access: Court Information, Privacy, and Technology, in* FUTURE TRENDS IN STATE COURTS 2011, at 11 (Carol R. Flango et al. eds., 2011) (reviewing the trends and issues relating to "an environment where most court systems maintain all or part of their information electronically").

2. A wide range of technical and policy challenges continue to be a part of the effort to increase public access to court records, both in the context of electronic filing and in the digitization of older court records. But one of the central issues is privacy.

3. The terms "private" and "sensitive" in the context of personally identifiable information are not necessarily coterminous. As we discuss in Part III, there are no uniform definitions for these terms and their scope is widely debated by privacy scholars. For readability, we use "sensitive information" to refer to all types of personally identifiable information that might raise privacy concerns.

4. Brief for the State, State v. Bright, 505 S.E.2d 317 (N.C. Ct. App. 1998), *disc. review allowed*, 525 S.E.2d 179 (N.C. 1998), 2012 WL 6685334 (also submitted in full to the North Carolina Supreme Court).

5. Notice of Appeal and Petition for Discretionary Review, Dean v. Cone Mills Corp., 322 S.E.2d 771 (N.C. 1984).

6. One important study of court records was conducted by Carl Malamud using automated software to search a large sample of court filings downloaded from the federal court's Public Access to Court Electronic Records (PACER) system. Malamud reported to the federal courts that a significant number of social security numbers and other types

The lack of empirical data hamstrings courts and archivists who are attempting to balance privacy interests with the public's right of access, as well as scholars looking to adapt privacy law and First Amendment doctrines to deal with the flood of public records going online.

This Article helps to fill this gap in our knowledge by analyzing a large corpus of court records from the North Carolina Supreme Court. These records are held by the Kathrine R. Everett Law Library at the University of North Carolina School of Law ("UNC Law Library"), one of several libraries with copies of briefs and court filings submitted to the North Carolina Supreme Court. Through sampling and content coding of briefs and appendices filed in cases decided between 1984 and 2000, we cataloged the types of sensitive information that appeared in these records,[7] determined the frequency of appearance of this information, and analyzed the context in which it appeared.

As is the case for courts and archivists everywhere, UNC Law Library personnel are grappling with the question of whether—and if so, how—to limit access to sensitive information that could cause financial, reputational, or emotional harm to individuals identified in the records. Although a patchwork of court rules, statutes, and government regulations provides some guidance as to the various categories of information that might raise privacy concerns, no studies address how often this information is likely to appear in court records, in what types of documents, and the specific context of its appearance.

The results of our research are valuable for a number of reasons. First, this study provides a highly granular view of sensitive information in judicial records. Based on a survey of the laws that apply to court records

of sensitive information were present in the downloaded files. *See* John Schwartz, *An Effort to Upgrade a Court Archive System to Free and Easy*, N.Y. TIMES (Feb. 12, 2009), http://www.nytimes.com/2009/02/13/us/13records.html (interviewing Malamud who states that he found several types of sensitive information in the downloaded case files); *see also* Letter from Carl Malamud to The Honorable Lee H. Rosenthal, Chair, Committee on Rules of Practice and Procedure, Judicial Conference of the United States (Oct. 24, 2008), https://public.resource.org/scribd/7512583.pdf (documenting 1,669 unredacted social security numbers and other proximate sensitive information in records from 32 district courts). Computer scientist Timothy Lee followed Malamud's report with a study that found some documents submitted to courts with intended redactions were not successfully redacted. Timothy B. Lee, *Studying the Frequency of Redaction Failures in PACER*, FREEDOM TO TINKER (May 25, 2011), https://freedom-to-tinker.com/blog/tblee/studying-frequency-redaction-failures-pacer. Our study provides a similar, but more extensive, examination of North Carolina Supreme Court briefs and appendices.

7. We did not code for the appearance of sensitive information in the court's decisions themselves.

as well as other privacy laws and scholarship, we have identified 140 types of sensitive information that may exist in court records. By coding for these information types, we are able to determine whether their frequency of appearance is correlated with case type, document type, or other contextual factors.

Second, an understanding of the types and context of sensitive information in the North Carolina Supreme Court's case files will help policymakers and judges evaluate the potential harms to privacy interests that might arise from the disclosure of sensitive information in court briefs and related records. Although this project examines only briefs and appendices filed in the North Carolina Supreme Court during a seventeen-year period, we expect that the results of our study will be generalizable to appellate court filings in many courts throughout the United States.

Third, this research will have practical implications for court personnel and archivists as they develop rules and practices for electronic filing of court records or the digitization of older records. Based on our informal survey of archivists and law librarians around the country, we have found that digitization initiatives are proceeding without a clear or consistent strategy for addressing privacy concerns. This project will help to identify and support the development of best practices for courts and archivists developing or implementing redaction protocols or making other choices regarding access and privacy.[8]

Finally, this research will be valuable to privacy scholars who can use our taxonomy and data to ground their normative arguments. Legal scholars, archivists, and law librarians have written extensively about the competing interests of government transparency and personal privacy. A number of these publications focus on court records, but little research provides empirical data concerning the frequency of sensitive information in particular types of court records. Although federal court rules require that some categories of information be redacted from court filings

8. All providers of court records face issues with regard to quality assurance, including data entry errors and other incorrect information. Courts and other archives that redact information face significantly greater quality assurance challenges. The results of this study should help to improve accuracy and reduce the cost of implementing redaction protocols, whether they are done manually (as the majority of such protocols are handled today) or through software. *See* Eric O. Scott et al., *Text Mining for Quality Control of Court Records* (Mitre Corp., Case Number 14-2510), http://mason.gmu.edu/~escott8/publications/Scott%20et%20al.,%20Text%20Mining%20for%20Quality%20Control%20of%20Court%20Records.pdf (presented at SemADoc 2014: Semantic Analysis of Documents Workshop, 16 September 2014).

submitted via electronic filing,[9] and some states are following in the same direction,[10] debate continues about how to address transparency and privacy in the context of e-filing systems and the digitization of older court records. This project supports discussion and policy shaping in these developing areas.

The loss of "practical obscurity" lies at the heart of the debate about privacy risks from online access to court records. Although court records have historically been available to the public for review, the information in these records was for practical purposes obscure because the records were "stored in such an inaccessible fashion that only the determined and resourceful could obtain them."[11] Peter Winn was one of the first to examine the loss of practical obscurity with the advent of electronic filing systems, noting that online access provided significant public benefits but raised serious privacy challenges.[12] Helen Nissenbaum and her coauthors

9. Federal e-filing rules now require redaction of several categories of data, including social security numbers and taxpayer numbers, dates of birth, names of minor children, financial account numbers, and in criminal cases, home address. *See* FED. R. APP. P. 25(a)(5); FED. R. CIV. P. 5.2; FED. R. CRIM. P. 49.1; FED. R. BANKR. P. 9037.

10. For example, since 2009 North Carolina has required e-filers to "exclude or partially describe sensitive, personal or identifying information such as any social security, employer taxpayer identification, driver's license, state identification, passport, checking account, savings account, credit card, or debit card number, or personal identification (PIN) code or passwords from documents filed with the court. In addition, minors may be identified by initials, and, unless otherwise required by law, social security numbers may be identified by the last four numbers." *Second Supplemental Rules of Practice and Procedure for the North Carolina eFiling Pilot Project*, *Rule 6.3*, N.C. COURT INFO. SYS. (Aug. 27, 2013), https://www.efiling.nccourts.org/manual/fiCourtRules.htm [hereinafter *N.C. eFiling Rules*] (referring to N.C.G.S. 132-1.10(d)). North Carolina also provides for non-parties to litigation to request removal or redaction of court documents available online for public viewing "if the document contains sensitive, personal or identifying information about the requester." *Id.* at Rule 6.4 (referring to N.C.G.S. § 132-1.10(f)). An excellent review of access policies for South Dakota state court records was produced in 2005. *See* Lynn E. Sudbeck, *Placing Court Records Online: Balancing Judicial Accountability with Public Trust and Confidence*, INSTITUTE FOR COURT MANAGEMENT, COURT EXECUTIVE DEVELOPMENT PROGRAM, PHASE III PROJECT (May 2005), https://www.ncsc.org/~/media/Files/PDF/Education%20and%20Careers/CEDP%20Papers/2005/SudbeckLynnCEDPFinal32905.ashx (recommending full access for the courts and litigants and an electronic version for the public with sensitive information redacted); *see also* D. R. Jones, *Protecting the Treasure: An Assessment of State Court Rules and Policies for Access to Online Civil Court Records*, 61 DRAKE L. REV. 375 (2013).

11. *Practical Obscurity*, SOC. OF AM. ARCHIVISTS, http://www2.archivists.org/glossary/terms/p/practical-obscurity (last visited Feb. 1, 2016); *see also* Woodrow Hartzog & Frederic Stutzman, *The Case for Online Obscurity*, 101 CALIF. L. REV. 1, 4–8 (2013).

12. *See* Peter A. Winn, *Online Court Records: Balancing Judicial Accountability and Privacy in an Age of Electronic Information*, 79 WASH. L. REV. 307 (2004) [hereinafter Winn, *Online Court Records*]; Peter A. Winn, *Judicial Information Management in an*

also have highlighted the chasm of difference between traditional in-person public access to court records at the courthouse and Internet access through Google and electronic filing systems.[13] Nissenbaum also articulated the importance of protecting information-flow norms threatened by disruptive technologies and practices such as the movement to make court records electronic and widely accessible.[14] Others too have explored this tension of interests,[15] and many propose that various types of information should be protected from broad exposure.[16]

Archivists and law librarians also have noted the challenges posed by electronic court records and documented their own efforts to address privacy concerns while developing new projects to digitize court documents in order to expand and facilitate public access.[17] For example,

Electronic Age: Old Standards, New Challenges, 3 FED. CTS. L. REV. 135 (2009) [hereinafter Winn, *Judicial Information Management*].

13. Amanda Conley, Anupam Datta, Helen Nissenbaum & Divya Sharma, *Sustaining Privacy and Open Justice in the Transition to Online Court Records: A Multidisciplinary Inquiry*, 71 MD. L. REV. 772 (2012).

14. Helen Nissenbaum, *Privacy as Contextual Integrity*, 79 WASH. L. REV. 119 (2004) (concluding that social and practical norms regarding information flow are superior to formal bifurcations of information into categories of public or private).

15. *See, e.g.*, Lynn M. LoPucki, *Court-System Transparency*, 94 IOWA L. REV. 481, 537 (2009) (suggesting privacy objections to highly transparent electronic federal court records should be addressed through removal of sensitive data and selective sealing of records and should not be used to shield the courts from scrutiny); Peter W. Martin, *Online Access to Court Records: From Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 855, 882–84 (2008) (warning that litigants are unlikely to adapt quickly to protect privacy, especially on behalf of non-litigants whose data may be in court filings); Will Thomas DeVries, *Protecting Privacy in the Digital Age*, 18 BERKELEY TECH. L.J. 283, 301 (2003) ("Digital technology is turning the asset of open government into a privacy nightmare. In the analog age, public records were all available, but languished in 'practical obscurity' in courthouse basements or isolated file cabinets.").

16. *See, e.g.*, Natalie Gomez-Velez, *Internet Access to Court Records*, 51 LOYOLA L. REV. 365 (2005) (recommending the use of protective orders and sealing to remove from public view high risk data elements and describing some states' decision to exclude categories of court records from online systems); Caren Myers Morrison, *Privacy, Accountability, and the Cooperating Defendant: Towards a New Role for Internet Access to Court Records*, 62 VAND. L. REV. 921, 969–78 (2009) (recommending redaction of names of cooperating defendants and other informants while increasing transparency in the use of these law enforcement practices); Kristin A. Henderson, *Lessons from Bankruptcy Court Public Records*, 23 LEGAL REFERENCE SERVS. Q. 55, 73, 76–77 (2004) (evaluating the sensitivity of information provided in bankruptcy proceedings and supporting the American Association of Law Libraries' proposal for redaction of sensitive information from bankruptcy court records accessible to the public through electronic case files).

17. *See, e.g.*, Michael Whiteman, *Appellate Court Briefs on the Web: Electronic Dynamos or Legal Quagmire?* 97 L. LIBRARY J. 467, 470–77 (2005) (describing the need to preserve access to court records and discussing the challenges associated with protecting privacy, including the Northern Kentucky Law Library's decision to refrain

the Montana State Law Library, which began scanning and posting Montana Supreme Court opinions and briefs in 1996, removed records it had already posted online to extract exhibits and appendices because the library found that these records contained a variety of sensitive information; the library ultimately reposted only the briefs with redactions.[18]

The goal of this study is to inform these scholarly and policy discussions about the appropriate balance between public access and privacy in the context of court records. We begin in Part II by noting that court records present a special challenge for privacy advocates. Unlike in many other areas of privacy law, court records are presumptively open to the public. Part II describes the origin of the right of public access to court records and examines its scope under the federal Constitution, common law, statutory law, and court rules. As we note in Part II, not all repositories of court records are obligated by law to provide public access. For many librarians and archivists, the question is not what the law requires, but rather what is the best approach for ensuring the protection of privacy interests while at the same time informing the public about the functioning of the court system.

In Part III, we survey privacy laws and privacy scholarship to create a taxonomy of sensitive information in court records.[19] Based on this survey,

from scanning appendices to briefs filed in the Kentucky Supreme Court). Archivists have been dealing with privacy concerns in a broad range of materials for many years and are considering how born-digital and digitized materials can be managed to address access and privacy. *See, e.g.*, Christopher A. Lee & Kam Woods, *Automated Redaction of Private and Personal Data in Collections: Toward Responsible Stewardship of Digital Heritage*, *in* PROCEEDINGS OF THE MEMORY OF THE WORLD IN THE DIGITAL AGE: DIGITIZATION AND PRESERVATION: AN INTERNATIONAL CONFERENCE ON PERMANENT ACCESS TO DIGITAL DOCUMENTARY HERITAGE (2012), http://ils.unc.edu/callee/p298-lee.pdf.

18. *See* Tammy A. Hinderman, *State Law Library Gets a 21st Century Makeover*, 32 MONT. L. REV. 6 (2007). According to Montana State Law Library reference librarian Tammy Hinderman, the library began providing online access to the Montana Supreme Court records in 2006, before realizing the records contained sensitive information that could facilitate identity theft and other privacy harms. *Id.* at 7. She explains that the library then removed all exhibits and appendices from the electronic version of the documents and redacted some information from the briefs before reposting them to the Internet. *Id.* In Kentucky, the Chase College of Law Library of Northern Kentucky University began its scanning of briefs from the state's supreme court by omitting appendices, both to address privacy concerns and to limit the burden on the library. *See* Whiteman, *supra* note 17, at 477.

19. Taxonomies of sensitive information appear throughout the law of the United States and other jurisdictions, and electronic filing systems at both the federal and state

we identified 140 types of sensitive information that might appear in court records and grouped them into thirteen categories. Part III explains the justifications—and shortcomings[20]—of our taxonomic approach and describes the sensitive information types we coded for in this project. Of course, not everyone will agree with our taxonomy. That is to be expected, given that privacy is itself a contested concept. Nevertheless, the taxonomy has proven to be helpful to us in the identification of the privacy risks that can arise from the public disclosure of court records at a time when privacy laws have limited or unclear application to such records. Moreover, we think our extensive taxonomy will be useful to others who wish to understand the broad range of privacy interests implicated by public records.

In Part IV, we provide an overview of our study design and methods. In short, we analyzed a stratified random sample of 504 court documents pulled from the briefs and other filings submitted to the North Carolina Supreme Court from 1984 to 2000. After performing content coding of the documents, we determined the frequency of appearance of each sensitive information type and identified relationships, patterns, and correlations between different information types and other coded variables, including trends over time.

In Part V, we present a summary of our findings. We begin by providing descriptive statistical information about the court records in our sample and the sensitive information they contain. We then examine the extent to which different types of sensitive information are related to various case and document characteristics. Although we suggest ways in which our data can aid in the assessment of the privacy risks that might arise from public access to court records, it is not our aim to tell courts or archivists what information, if any, should be redacted or what documents should be withheld from online access or otherwise managed for privacy protection.[21]

Instead, in Part VI, we discuss how our study can inform the debate about privacy and court records and how our results can help to identify and remedy some of the challenges courts and archivists are likely to face if they decide to implement procedures for addressing privacy concerns in

level rely extensively on predefined lists of information types that must be handled with special care. *See infra* notes 89–98 and accompanying text.

20. Scholars have long criticized this approach because it relies on debated definitions of privacy, ignores contextual variations in privacy, and presents implementation challenges because of these definitional and contextual problems. We discuss these concerns and how we dealt with them in Part III.

21. We plan to address these normative questions in subsequent articles.

court records. We made several important findings in this regard. First, although a wide variety of sensitive information appears in the court records we sampled, it is not uniformly distributed throughout the records. Most of the documents contained relatively few incidences of sensitive information while a handful of documents contained a large number of pieces of sensitive information. Second, we found that court records vary substantially in the types and frequency of sensitive information they contain. Sensitive information in seven categories—"Location," "Identity," "Criminal Proceedings," "Health," "Assets," "Financial Information," and "Civil Proceedings"—appeared much more frequently than information in the other categories we identified. Third, we found that information associated with criminal proceedings, such as witness and crime victim names, is pervasive in court records, appearing in all types of cases and records. Information in the "Criminal Proceedings" category not only appeared in most of the documents we reviewed, but also appeared more often in those documents than any other category of sensitive information. Fourth, the data showed that criminal cases have disproportionately more sensitive information than civil or juvenile cases., with death penalty cases far exceeding all other case types. Fifth, we found that appendices are generally not quantitatively different than legal briefs in terms of the frequency and types of sensitive information they contain, a finding that goes against the intuition of many privacy advocates. Sixth, we saw no overarching trends in the frequency of sensitive information during the seventeen-year period under study.

We close by providing some suggestions for courts and archivists seeking to manage sensitive information in court records. A number of practices have been introduced or recommended, including redaction of electronic records, redaction of both electronic and print records, removal of categories of court records from Internet access, and increased filing of court documents under seal. Our research will help courts and archivists evaluate these approaches.

II. PUBLIC ACCESS TO COURTS AND COURT RECORDS

A. THE RIGHT TO ACCESS COURT PROCEEDINGS AND RECORDS

Public access to the courts has a long and venerated history in America, even predating enactment of the United States Constitution.[22]

22. *See, e.g.*, Leucadia, Inc. v. Applied Extrusion Techs., Inc., 998 F.2d 157, 161 (3d Cir. 1993) (concluding that "[t]he existence of this right, which antedates the Constitution and which is applicable in both criminal and civil cases, is now 'beyond

This openness serves many salutary functions, including ensuring that our system of justice functions fairly and is accountable to the public.[23] As Chief Justice Warren Burger noted in *Richmond Newspapers, Inc. v. Virginia*:

> The early history of open trials in part reflects the widespread acknowledgment, long before there were behavioral scientists, that public trials had significant community therapeutic value. Even without such experts to frame the concept in words, people sensed from experience and observation that, especially in the administration of criminal justice, the means used to achieve justice must have the support derived from public acceptance of both the process and its results.[24]

Public access also has been extended to many of the records associated with court proceedings.[25] Access to judicial records plays a critical role in fostering public awareness about the operation of the courts because so few people are able to attend court proceedings in person and because most courts do not generally allow live or archival recordings. The movement by courts and archivists to allow online access to court records has made it possible for many more people to stay informed about the functioning of the judicial system.[26] Online access also has a leveraging effect because it makes it possible for the media to cover court proceedings at a lower cost and allows for greater depth of reporting at a time when many media

dispute.'"); Winn, *Online Court Records*, *supra* note 12, at 307 (noting that "the legal system has inherited from the Enlightenment a presumption of openness"); Conley et al., *supra* note 13, at 785 (observing that "the right to open courts and their records is actually as longstanding as our right to the courts and to justice itself").

23. *See* Globe Newspaper Co. v. Super. Ct., 457 U.S. 596, 606 (1982) ("Public scrutiny of a criminal trial enhances the quality and safeguards the integrity of the factfinding process, with benefits to both the defendant and to society as a whole.").

24. Richmond Newspapers, Inc. v. Virginia, 448 U.S. 555, 570–71 (1980).

25. The public's right of access to court records is not absolute and may be restricted in some circumstances. *See infra* notes 30–33 and accompanying text.

26. *See, e.g.*, Lynn E. Sudbeck, *Placing Court Records Online: Balancing Judicial Accountability with Public Trust and Confidence: An Analysis of State Court Electronic Access Policies and a Proposal for South Dakota Court Records*, 51 S.D. L. REV. 81, 91 (2006) (noting that a "frequently mentioned benefit" of electronic access to court records is that it responds "to the needs of South Dakota's rural court users, that is, [it] 'levels the geographic playing field' by allowing persons located in great distances from the courthouse to access public information" (citation omitted)). Online access to court records also allows litigants, lawyers, and educators to scrutinize legal strategy and rhetoric. *See* Anna P. Hemingway, *Making Effective Use of Practitioners' Briefs in the Law School Curriculum*, 22 ST. THOMAS L. REV. 417 (2010) (advocating use of practitioners' briefs to teach persuasive writing and legal analysis).

organizations are cutting back on the number of reporters assigned full-time to the courts.[27]

Although public access to court records is longstanding and deeply ingrained in our legal system, courts can—and often do—impose limits on public access. The First Amendment provides a right of access to court proceedings and to many records,[28] as does federal and state common law.[29] These rights, however, are not absolute.[30] While the precise standard that a court must apply will vary depending on the source of the public's right of access, in general courts must at least conclude that the interest in prohibiting disclosure outweighs the strong presumption of public access. In *Nixon v. Warner Communications*, for example, the United States Supreme Court instructed that "[e]very court has supervisory power over its own records and files" and that the federal

27. *See Panel One: General Discussion on Privacy and Public Access to Court Files*, 79 FORDHAM L. REV. 1, 13 (2010) (quoting testimony of Lucy Dalglish before the Privacy Subcommittee of the Judicial Conference Standing Committee on the Federal Rules).

28. *See, e.g.*, *Richmond Newspapers*, 448 U.S. at 575 (finding First Amendment right of public access to criminal trials and noting that "[i]n guaranteeing freedoms such as those of speech and press, the First Amendment can be read as protecting the right of everyone to attend trials so as to give meaning to those explicit guarantees"). Although the U.S. Supreme Court has not explicitly held that a First Amendment right of access applies in civil cases, most of the federal circuits that have addressed this issue have recognized such a right. *See, e.g.*, Westmoreland v. Columbia Broad. Sys., Inc., 752 F.2d 16, 23 (2d Cir. 1984); Publicker Indus., Inc. v. Cohen, 733 F.2d 1059, 1067–71 (3d Cir. 1984). Courts have also applied a constitutional right of access to the judicial records associated with criminal and civil proceedings. *See, e.g.*, Associated Press v. United States Dist. Court. for Cent. Dist. of Cal., 705 F.2d 1143, 1145 (9th Cir. 1983); *In re* Search Warrant, 855 F.2d 569, 573 (8th Cir. 1988); Newsday LLC v. Cnty. of Nassau, 730 F.3d 156, 164 (2d Cir. 2013); *Publicker Indus.*, 733 F.2d at 1074. *But see* Zenith Radio Corp. v. Matsushita Elec. Indus. Co., 529 F. Supp. 866, 908 (E.D. Pa. 1981) ("With respect to the question whether the common law right to inspect and copy [discovery materials] has a constitutional dimension, we conclude that it does not.").

29. *See* Nixon v. Warner Commc'ns, 435 U.S. 589, 597 (1978) (recognizing a federal common law right to "inspect and copy public records and documents, including judicial records and documents"); Richard J. Peltz et al., *The Arkansas Proposal on Access to Court Records: Upgrading the Common Law with Electronic Freedom of Information Norms*, 59 ARK. L. REV. 555, 591–94 (2006) (discussing various state approaches).

30. When the right of public access arises under the First Amendment, "it must be shown that the denial [of access] is necessitated by a compelling governmental interest, and is narrowly tailored to serve that interest." Globe Newspaper Co. v. Super. Ct., 457 U.S. 596, 607 (1982). When the right of access is merely a common right, courts have more leeway in denying access and can balance the presumption of public access against other interests, including the possibility of prejudicial pretrial publicity; the danger of impairing law enforcement or judicial efficiency; and the protection of the legitimate privacy interests of litigants and other trial participants, such as witnesses, victims, and jurors. *See, e.g.*, *Nixon*, 435 U.S. at 598.

common law right of access could be denied when "court files might . . . become a vehicle for improper purposes."[31] Among the improper purposes the Court noted were uses "to gratify private spite or promote public scandal," as "reservoirs of libelous statements for press consumption," and as a source of unfair competitive "business information."[32] Although there is considerable variation among the states, state common law rights of access also are typically qualified rights, allowing courts to restrict public access if an overriding interest supports closure or sealing of specific information. In California, for example, court records are "presumptively open to the public and [court proceedings and records] should not be closed except for compelling countervailing reasons."[33]

In addition to constitutional and common law rights of access, a number of state and federal statutes also provide a public right of access to court records. At the federal level, access to court records is governed by rules and policies promulgated by the Administrative Office of the U.S. Courts on behalf of the federal judiciary pursuant to the Rules Enabling Act.[34] At the state level, a variety of statutory authority provides for and impacts public access to judicial records. For example, every state has a public records statute, although not all of these statutes explicitly address access to court records.[35] In those states that do have a public records law that covers judicial records, rights of access are typically governed by both the statute and court rules.[36]

31. *Nixon*, 435 U.S. at 598.

32. *Id.*

33. Pantos v. City & Cnty. of S.F., 198 Cal. Rptr. 489, 492 (Cal. Ct. App. 1984) (citations omitted).

34. 28 U.S.C. §§ 2071–2077. The Rules Enabling Act authorizes the Supreme Court to prescribe general rules of practice and procedure and rules of evidence for the federal courts. Pursuant to Section 2073 of the Rules Enabling Act, the U.S. Judicial Conference has established procedures to govern the work of its Standing Committee and its advisory rules committees. *See* UNITED STATES COURTS, CRIMINAL JUSTICE ACT GUIDELINES, GUIDE TO JUDICIARY POLICY § 440 (2014), http://www.uscourts.gov/file/2932/download.

35. *See* REPORTERS COMMITTEE FOR FREEDOM OF THE PRESS, OPEN GOVERNMENT GUIDE (5th ed. 2006) (providing summaries of public records laws in all U.S. jurisdictions), http://www.rcfp.org/ogg/index.php. When a state's public records law is silent, the state's highest court may define the scope and procedures for public access.

36. *See, e.g.*, Anderson v. Home Ins. Co., 924 P.2d 1123, 1126 (Colo. App. 1996) (interpreting the state's public records law and court rules to hold that there is a strong presumption that all court records are open); Doe v. New York Univ., 786 N.Y.S.2d 892, 899 (N.Y. Sup. Ct. 2004) (applying both statutory law and court rules).

As a result, many states have multiple overlapping sources of law that require—and potentially limit—public access.[37] In North Carolina, which is illustrative of the law in many states, public access is governed by, *inter alia*, a common law right of access,[38] a constitutional right of access rooted in both the First Amendment to the U.S. Constitution and article 1, section 18 of the N.C. Constitution, which states that "[a]ll courts shall be open,"[39] and court rules that specify how court records are to be handled, including rules for electronic-filing.[40]

Furthermore, the North Carolina General Assembly, through the state's public records law ("NC PRL") and other statutes, has both expanded and narrowed the public's right of access.[41] The NC PRL, which states that all state records "are the property of the people," is applicable to every agency of the North Carolina government, including the judiciary.[42]

---

37. *See* Richard J. Peltz, et al., *The Arkansas Proposal on Access to Court Records: Upgrading the Common Law with Electronic Freedom of Information Norms*, 59 ARK. L. REV. 555, 591 (2006) (noting that "[s]ome states decided that only one type of law was necessary to adequately provide a right of access, while others applied multiple types of law to provide more depth to their access law").

38. *See* Virmani v. Presbyterian Health Servs. Corp., 515 S.E.2d 675, 691 (N.C. 1999) (observing that "[a]t least since 1887, this Court has recognized a common law right of the public to inspect public records"). As with the federal common law, the common law right of access in North Carolina is a qualified right. The decision to deny access "is left to the sound discretion of the trial courts, a discretion to be exercised in light of the relevant facts and circumstances of the particular case." *In re* Investigation into Death of Cooper, 683 S.E.2d 418, 425 (N.C. App. 2009) (internal quotation marks omitted).

39. *Virmani*, 515 S.E.2d at 692 (holding that the N.C. Constitution guarantees a qualified constitutional right on the part of the public to attend civil court proceedings and access court records). In the words of the North Carolina Supreme Court: "That courts are open is one of the sources of their greatest strength." Raper v. Berrier, 97 S.E.2d 782, 784 (N.C. 1957).

40. For example, N.C.'s eFiling Rule 6.3 states, in part:

> Except where otherwise expressly required by law, filers must comply with G.S. 132-1.10(d) to exclude or partially describe sensitive, personal or identifying information such as any social security, employer taxpayer identification, driver's license, state identification, passport, checking account, savings account, credit card, or debit card number, or personal identification (PIN) code or passwords from documents filed with the court.

N.C. eFiling Rules, *supra* note 10, at Rule 6.3.

41. *See In re* Investigation into Death of Cooper, 683 S.E.2d at 425; N.C. GEN. STAT. § 132-1 (2015).

42. N.C. GEN. STAT. § 132-1(b) (2015) ("The public records and public information compiled by the agencies of North Carolina government or its subdivisions are the property of the people. Therefore, it is the policy of this State that the people may

The NC PRL does not grant court records any special dispensation from public access requirements, except to define two narrow exceptions to the law. The first exception allows for the withholding of settlements in medical malpractice actions against public hospitals.[43] The other exception makes arrest and search warrants confidential until they have been returned.[44] Various statutes outside the NC PRL also treat some court documents as confidential. These include records of grand jury proceedings;[45] most adoption records;[46] and reports of cases of juvenile abuse, neglect, or dependency.[47] Other than these significant exceptions, almost all court records are subject to public inspection under the NC PRL unless otherwise specifically restricted by law.[48]

Given this overlapping and sometimes ambiguous legal authority, it should come as no surprise that individual judges and court clerks frequently struggle with how to implement the public's right of access to court records. As Amanda Conley, Anupam Datta, Helen Nissenbaum, and Divya Sharma note, "restrictions on access trickle down from state and federal appellate courts to the local courthouses themselves, where state and local law, custom, and in some cases simply the whims of court clerks determine which information in the court record will actually be made available to the public, and how."[49]

Moreover, librarians and archivists, who may not be bound by law to provide public access to court records,[50] have an even broader range of

obtain copies of their public records and public information free or at minimal cost unless otherwise specifically provided by law.").

43. N.C. GEN. STAT. § 132-1.3(a) (2015).

44. N.C. GEN. STAT. § 132-1.4(k) (2015).

45. N.C. GEN. STAT. § 15A-623 (2015).

46. N.C. GEN. STAT. § 48-9-102 (2015).

47. N.C. GEN. STAT. § 7B-2901 (2015).

48. *See* News & Observer Pub. Co. v. Poole, 412 S.E.2d 7, 19 (N.C. 1992) ("[W]e hold that in the absence of clear statutory exemption or exception, documents falling within the definition of 'public records' in the Public Records Act must be made available for public inspection.").

49. Conley et al., *supra* note 13, at 787.

50. The applicability of public records statutes to publicly supported libraries' collections is not well established. Public libraries have been described as requiring autonomy to add and withdraw materials from their collections, at least in the context of First Amendment analysis. *See* United States v. Am. Library Ass'n, 539 U.S. 194, 195 (2003) (Rehnquist, C.J., plurality opinion) ("To fulfill their traditional missions of facilitating learning and cultural enrichment, public libraries must have broad discretion to decide what material to provide to their patrons."). State archives, however, tend to have statutory requirements for providing access to public records. *See* Carol D. Billings, *State Government Efforts to Preserve Electronic Legal Information*, 96 L. LIBRARY J. 625, 626 (2004) (noting that "most state libraries that operate the depository programs and

options for dealing with sensitive information in court records. For many, the question is not what the law requires, but rather what is the best policy for ensuring the protection of privacy interests while at the same time informing the public about the functioning of the court system.[51] As a result, some libraries exclude whole categories of records from public access,[52] whereas others engage in targeted redactions of sensitive information based either on their own assessment of what is private[53] or on the electronic filing rules adopted by their courts.[54] Alternatively, several libraries have adopted a middle-ground approach. They provide mediated access to court records, allowing only bibliographic information to be discoverable on the Internet, not the contents of the records themselves,[55] or limiting access to unaltered briefs to registered library users.[56]

state archives with responsibility for preserving records lack rule-making and enforcement powers to require compliance").

51. *See, e.g.*, Hinderman, *supra* note 18, at 7 (discussing the Montana State Law Library's efforts to balance privacy and public access concerns); Whiteman, *supra* note 17, at 477 (describing the approach taken by Northern Kentucky University's law library). Even if a library or other archive decides to make case files available without any restrictions on access, it should not face legal liability if the records contain information that violates privacy law. *See* Cox Broad. Corp. v. Cohn, 420 U.S. 469, 496 (1975) ("Once true information is disclosed in public court documents open to public inspection, the press cannot be sanctioned for publishing it.").

52. *See* Hinderman, *supra* note 18, at 7 (describing the exclusion of appendices); Whiteman, *supra* note 17, at 477 (same).

53. The Montana State Law Library redacted social security numbers, dates of birth and other "obviously private information" from the briefs in its database of supreme court briefs. Hinderman, *supra* note 18, at 7. The Blakley Law Library at the Arizona State University Sandra Day O'Connor College of Law has posted digital versions of state appellate and supreme court briefs to the Internet with the caveat, "Certain types of personal information may have been removed from briefs on the Arizona Memory Project to allow for online publication." *About Collection*, ARIZONA MEMORY PROJECT, http://azmemory.azlibrary.gov/cdm/landingpage/collection/asuross (last visited May 31, 2015).

54. *See* Faye Jones & Caroline Osborne, *Lessons Learned: Creating Digital Collections and Privacy: Best Practices*, Presentation at the Southeastern Association of Law Libraries Annual Meeting (April 16, 2015) (presentation slides on file with authors) (comments of Faye Jones, describing the Florida State University College of Law Library's collaboration with other Florida law libraries to provide Internet access to state supreme court briefs and citing Florida public records laws (FLA. STAT. §§ 119.01–.15 as well as FLA. R. JUD. ADM. 2.420), which outline confidentiality guidelines for filing of court records including redaction).

55. *Id.* (comments of Caroline Osborne, explaining Washington and Lee Law Library's project to digitize and not redact copies of Virginia Supreme Court briefs, to store the digital briefs in a "dark archive," and to develop policies and procedures for responding to requests for individual briefs, citing state statutes on freedom of information (VA. CODE §§ 2.2-3700–3714), prohibition of posting certain information

B. COUNTERVAILING INTERESTS

Court records contain a variety of information that can cause harm to individual, organizational, and governmental interests. A court's file for a single case may consist of thousands of documents, including motions, pleadings, briefs, transcripts, exhibits entered into evidence, and records and responses produced during pre-trial discovery that have been filed with the court.[57] For individuals, information ranging from social security numbers to sexual history can appear in these documents raising, among other concerns, the risk of identity theft and reputational harm.[58] For businesses and other organizations, court records can contain trade secrets and other confidential information.[59] For the government, information in court records such as the names of confidential informants and descriptions of intelligence gathering techniques can potentially harm national security or undermine law enforcement efforts.[60] Although all of these countervailing interests are worthy of study, our focus is on the impact that the disclosure of sensitive information in court records can have on individuals.

Given that "[t]he courts are a stage where many of life's dramas are performed, where people may be shamed, vindicated, compensated, punished, judged, or exposed,"[61] it is natural that court records, which

to the Internet (VA. CODE § 17.1-293), and personal information privacy (VA. CODE § 59.1.443.2)).

56. *See Policies for Utah Court Briefs*, HOWARD W. HUNTER LAW LIBRARY, J. RUBEN CLARK LAW SCHOOL, BRIGHAM YOUNG UNIVERSITY, http://digitalcommons.law.byu.edu/utah_court_briefs/policies.html (last visited Feb. 1, 2016) (explaining that briefs are "supplied to the Hunter Law Library by the courts for the purposes of legal scholarship and academic research. The Hunter Law Library provides this collection as authorized by the Utah Courts. The Law Library is not responsible for the selection or content of individual records.").

57. *See* Conley et al., *supra* note 13, at 781 (noting that "[e]ach and every form filled out by the parties, their lawyers, or by related third parties (witnesses, jurors, etc.) potentially contains vast amounts of personal data including home or school addresses, places of employment, birthdates, and, in many cases, Social Security numbers."). Some documents such as sealed discovery materials, *see* Leucadia, Inc. v. Applied Extrusion Techs, Inc., 998 F.2d 157, 163–65 (3d Cir. 1993), and certain financial information about the parties, *see* United States v. Lexin, 434 F. Supp. 2d 836, 849 (S.D. Cal. 2006), are often excluded from the public court record.

58. *See infra* Part III.

59. *See* Kyle J. Mendenhall, *Can You Keep A Secret? The Court's Role in Protecting Trade Secrets and Other Confidential Business Information from Disclosure in Litigation*, 62 DRAKE L. REV. 885 (2014).

60. *See* Laura K. Donohue, *The Shadow of State Secrets*, 159 U. PA. L. REV. 77, 78 (2010).

61. Conley et al., *supra* note 13, at 774.

serve as a chronicle of these dramas, are littered with private and sensitive information. In fact, they are full of information not just about the parties in a case, but also about witnesses, family members, victims, and jurors, among other individuals who are brought willingly or unwillingly into a legal dispute.

Although concerns about private information in court records existed long before the Internet, many commentators see the move to electronic court records as effectuating a qualitative shift in the balance between the competing interests of public access and individual privacy. Not so long ago it was difficult and time-consuming to access and search an entire case file. Today, with the advent of electronic court records and online access, it takes little effort to find and link information across cases, courts, and states. The following sections highlight the most pressing concerns that arise from the transition to online court records. We then dive much more deeply into these issues in Parts III and VI.

1. Privacy and the Loss of Practical Obscurity

Courts, like other institutions in our society, are in the midst of a transformation. The largely paper-based world of the twentieth century is giving way to an interconnected, electronic world where physical and temporal barriers to public access are evaporating. Over the past decade, courts across the country have been moving with alacrity to digitize their records and make them available to the public online.[62] Some courts are doing this work themselves, while others are relying on third parties, such as libraries and other archives, to make online access to historical records possible. A growing number of courts also require litigants to file their pleadings, motions, and other documents in electronic format.[63]

62. *See* John T. Matthias, *E-Filing Expansion in State, Local, and Federal Courts 2007*, *in* FUTURE TRENDS IN STATE COURTS 2007, at 34 (highlighting state courts' efforts to move to e-filing and the conversion of paper case documents into digital images), http://ncsc.contentdm.oclc.org/cdm/ref/collection/tech/id/570; HON. PAUL H. ANDERSON, FUTURE TRENDS IN PUBLIC ACCESS: COURT INFORMATION, PRIVACY, AND TECHNOLOGY 11 (2011) (reviewing the trends and issues relating to "an environment where most court systems maintain all or part of their information electronically").

63. *See, e.g.*, Peter W. Martin, *Online Access to Court Records—from Documents to Data, Particulars to Patterns*, 53 VILL. L. REV. 855, 872 (2008) ("By the end of 2007, electronic filing was an option in nearly all federal trial courts and was mandatory in a large number."); Eric J. Magnuson & Samuel A. Thumma, *Prospects and Problems Associated with Technological Change in Appellate Courts: Envisioning the Appeal of the Future*, 15 J. APP. PRAC. & PROCESS 111, 114 (2014) ("By late 2012, all federal courts of appeals were using electronic filing (e-filing)."); Matthias, *supra* note 62, at 34 (reporting

As discussed in the previous section, court records have for centuries been open for public review. Yet the difficulty of actually accessing individual records—for example, traveling to the courthouse, identifying the relevant case, finding the sought after document, and copying the information—made the information in these records practically obscure in the sense that private and sensitive information could remain in the records without creating a significant risk of harm. Today, this practical obscurity is vanishing. Although the specifics of electronic access vary by state (and sometimes by court), in most federal courts and many state jurisdictions anyone can access a court's electronic case database through a website interface.[64] That interface typically provides the ability to search by party names, case type, keywords, and other information, as well as providing case-by-case browsing. If users wish to copy a document, they can usually do so by downloading it as a PDF file.[65]

The loss of practical obscurity that has resulted from this nearly frictionless access to court records lies at the heart of the debate about the privacy risks arising from online access. The Supreme Court recognized the importance of practical obscurity in holding that rap sheets aggregating public—but difficult to assemble—information qualify for a privacy exemption from disclosure under the federal Freedom of Information Act.[66] The Court stated, "Plainly there is a vast difference between the public records that might be found after a diligent search of

that as of 2007, twenty-six states had adopted court rules enabling e-filing statewide or in at least one court).

64. Some courts charge for access, some merely require registration, while others do not require either payment or registration.

65. In jurisdictions that have public records laws that cover court records, a requester may even be entitled to a copy of a court's entire case database, though some limitations might apply. *See* LexisNexis Risk Data Mgmt. Inc. v. N.C. Admin. Office of Courts, 776 S.E.2d 651, 652 (N.C. 2015) (finding that the court's Automated Criminal/Infraction System (ACIS) database was a public record under the North Carolina Public Records Act subject to a limiting statutory provision requiring requesters to secure a nonexclusive contract and pay for reasonable cost recovery).

66. U.S. Dept. of Justice v. Reporters Committee for Freedom of the Press, 489 U.S. 749, 762–63 (1989) (describing the view that aggregated information provides no more privacy harm than its discrete components as a "cramped notion of personal privacy"). It should be noted that the Court's decision in *Reporters Committee for Freedom of the Press* did not address access to court records, but rather a request for access under FOIA to a database of criminal history information compiled by the FBI. *Id.* at 751–52. The standard for determining whether public access can be denied under FOIA is less demanding than the standard for restricting access to court records; all that the government was required to show was that disclosure "could reasonably be expected to constitute an unwarranted invasion of personal privacy." *Id.* at 756 (quoting 5 U.S.C. § 552(b)(7)(C)).

courthouse files, county archives, and local police stations throughout the country and a computerized summary located in a single clearinghouse of information."[67]

2. Navigating the Transition to Online Court Records

As a result of these and other concerns, court administrators, judges, lawyers, librarians, and legislators are in active discussion about how to navigate the transition to online court records.[68] Privacy scholars have also been trying to influence this transition. Indeed, a number of legal scholars consider practical obscurity to be a stand-in for privacy interests and now, with the loss of this obscurity, are suggesting that courts and archivists should implement various approaches to obscuring sensitive information in court records.[69] Other scholars also have explored the tension between privacy and public access to court records, with some recommending a substantial curtailment of public access through redaction of electronic and print records, restricted public access, removal of categories of court records from Internet access, and increased filing of court documents under seal.[70]

Although important theoretical work is being done with regard to the nature and extent of the privacy interests implicated by public access to court records,[71] we are only just beginning to develop a sufficient body of

67. *Id.* at 764.

68. See Conley et al., *supra* note 13, at 776 (noting that "public and internal deliberations over state access policies have remained actively in progress").

69. *See* Hartzog & Stutzman, *supra* note 11, at 41 ("Obscurity obligations would not aim to completely curtail information disclosure; rather, they would seek to minimize the likelihood of discovery, comprehension, or contextualization."); Steven C. Bennett *Pleadings, Privacy and Ethics: Protecting Privacy in Litigation Documents*, 2 REYNOLDS CT. & MEDIA L.J. 25 (2012); Daniel J. Solove, *Access and Aggregation: Public Records, Privacy and the Constitution*, 86 MINN. L. REV. 1137 (2002); Will T. DeVries, *supra* note 15.

70. *See, e.g.*, Gomez-Velez, *supra* note 16, at 431–32 (examining the decisions of some states to exclude categories of court records from online systems); Morrison, *supra* note 16, at 925–27 (recommending redaction of identifying information of cooperating defendants and other informants while increasing transparency in using these law enforcement practices); Henderson, *supra* note 16, at 76–77 (supporting the American Association of Law Libraries' advocacy for redaction of sensitive information from bankruptcy court records accessible to the public through electronic case files). Not all scholars argue for restricting public access. *See, e.g.*, Lynn LoPucki, *The Politics of Research Access to Court Data*, 80 TEX. L. REV. 2161 (2002) (arguing against selective restriction of access to court records to enable better empirical research about the courts).

71. *See, e.g.*, Nissenbaum, *supra* note 14, at 136–38 (concluding that accepted social and practical norms for information flows are superior to formal bifurcations of information into categories of public or private); Hartzog & Stutzman, *supra* note 11, at 3–4 (suggesting that the concept of "online obscurity" is a critical component of online

research that examines the risks to privacy when court records are made available through the Internet compared with long-standing public access that was practically obscure due to the logistical barriers to access.

Our present research helps to fill this gap in our knowledge. Empirical data about the frequency and context of sensitive information in the North Carolina Supreme Court's files will allow policymakers and scholars to better understand and evaluate the range of privacy risks that can arise from online court records.

III. A SENSITIVE INFORMATION TAXONOMY FOR COURT RECORDS

Our project draws on the longstanding approach to privacy of identifying certain types of information that present risks of harm that can be reduced through restrictions on public exposure. Taxonomies of sensitive information appear throughout the law of the United States and other jurisdictions, and electronic filing systems at both the federal and state level rely extensively on predefined lists of information types that must be handled with special care.[72] The use of sensitive information taxonomies is pervasive because they provide an attractive, seemingly simple solution, for balancing privacy and competing interests. Indeed, this approach to privacy is the basis of much of privacy law.[73]

privacy and developing an analytical framework for use by lawmakers and courts); Solove, *supra* note 69, at 1176–78 (criticizing the "secrecy paradigm" in privacy discourse and suggesting that there is an "expectation of limits on the degree of accessibility" to public records).

72. *See, e.g.*, HIPAA Privacy Rule, 45 C.F.R. § 164.514(b)(2) (2015) (listing seventeen specific identifiers to be removed to "de-identify" personal health information); Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data, art. 2, sec. (a), 1995 O.J. (L 281) 31, 38, http://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:31995L0046&from=en (defining personal data). As Paul Ohm has noted, although many scholars have turned away from the list-based approach to privacy protection, U.S. law remains largely grounded in this model. Paul Ohm, *Sensitive Information*, 88 S. CAL. L. REV. 1125, 1125–32 (2015).

73. *See* Ohm, *supra* note 72, at 1128–29 ("The great variety of regulations, law, technical standards, and corporate practices that have been implemented to protect the privacy of information stored in databases share at their core this unifying construct [of sensitive information.]").

A. THE CHALLENGES OF CREATING A TAXONOMY OF SENSITIVE INFORMATION

1. Building a Taxonomy on Debated Definitions of Privacy and Related Concepts

The creation of a comprehensive taxonomy of sensitive information types is a challenge because privacy law and policy are not grounded in a coherent understanding of or approach to privacy.[74] Moreover, some conceptions of privacy simply do not lend themselves to a sensitive-information approach. In addition, disagreement about the proper role of related concepts of confidentiality, practical obscurity, and contextual privacy increases the difficulty of creating a taxonomy of sensitive information.

One of the core problems is the lack of consensus about the underlying interests and risks that define privacy. Financial integrity,[75] personal safety,[76] non-discrimination,[77] confidential access to professional advice,[78]

74. Daniel J. Solove, *A Taxonomy of Privacy*, 154 U. PA. L. REV. 477, 477–78 (2006) ("Privacy is a concept in disarray. Nobody can articulate what it means. As one commentator has observed, privacy suffers from 'an embarrassment of meanings.'").

75. Identity-theft statutes prohibit the disclosure of data such as financial account numbers and PIN codes. *See, e.g.*, N.C. GEN. STAT. § 14-113.20(b) (2015); MASS. GEN. LAWS ch. 266, § 37E(a) (2015). Data security breach notification statutes require that companies and government entities encourage individuals to monitor their accounts for tampering if sensitive data is not kept confidential. N.C. GEN. STAT. § 75-65 (2015); CAL. CIV. CODE § 1798.82 (West 2015). Gramm-Leach-Bliley mandates notice requirements to allow bank customers to opt-out of permitted sharing of some of their financial information. Gramm-Leach-Bliley Act (GLBA), 15 U.S.C. § 6802 (2012).

76. Publication of location information can place persons, particularly police officers, cooperating defendants, and victims of stalking, in harm's way. Grayson Barber, *Personal Information in Government Records: Protecting the Public Interest in Privacy*, 25 ST. LOUIS U. PUB. L. REV. 63, 77 (2006); Morrison, *supra* note 16, at 971. Some federal and state statutes offer privacy protection under limited circumstances to particular groups. *See, e.g.*, Family Educational Rights and Privacy Act (FERPA), 20 U.S.C. § 1232g(a)(5)(B) (2012) (mandating the option for parents to opt-out from the publishing of student directory information); CAL. GOV. CODE § 6254.21 (West 2015) (prohibiting the posting of any elected or appointed official's home address or telephone number to the Internet without written permission).

77. Although Federal EEO law does not require non-disclosure of protected class status, Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e-2, some employers have developed best practices for avoiding related questions in order to provide a defense that they could not have based hiring decisions on information they did not have. In effect, privacy works as a barrier to discrimination. *See* Chad Derum & Karen Engle, *The Rise of the Personal Animosity Presumption in Title VII and the Return to "No Cause" Employment*, 81 TEX. L. REV. 1177 (2003) (arguing that a shift in the presumption of failure to hire for discriminatory reasons has occurred in favor of employers); Daniel J. Bugbee, *Employer's Beware: Violating USERRA through Improper Pre-Employment*

and protection of intellectual development space for certain creative pursuits and for children are some of the interests protected under the umbrella of privacy law.[79] Other concepts associated with privacy include autonomy, dignity, and liberty.[80] Some of these privacy interests relate to rights against the government while others address privacy in the context of private relationships. In addition, although privacy is generally considered a personal interest, it is also advanced as an important benefit to society.[81]

Another point of debate is the authority for defining privacy interests. Some approaches embrace the idea that privacy is a personal choice.[82]

Inquiries, 12 CHAP. L. REV. 279 (2008) (discussing pre-employment inquiries under the Uniformed Services Employment and Reemployment Rights Act which protects those who served in the military).

78. Evidentiary privileges such as the attorney-client privilege are designed to encourage disclosure by providing confidentiality. MODEL CODE OF PROF'L CONDUCT R. 1.6 (1983).

79. The Children's Online Privacy Protection Act (COPPA) provides protections for children in the online environment, including parental consent requirements before certain personal information can be collected from a child. Children's Online Privacy Protection Act of 1998, Pub. L. No. 105-277, 112 Stat. 2681-728 (codified at 15 U.S.C. §§ 6501–6506 (2012)). Intellectual privacy is protected in state library privacy statutes and in the case of California, an e-reader privacy statute. California Online Privacy Protection Act (CalOPPA), CAL. BUS. & PROF. CODE § 22575 (West 2015).

80. Protections of some areas of personal integrity are recognized as constitutional freedoms from government intrusion, and this body or bodies of law are characterized as decisional privacy, information privacy, and/or liberty protections. *See, e.g.*, Griswold v. Conn., 381 U.S. 479 (1965) (holding that the right of marital privacy was violated by a statute restricting the use of or provision of advice in support of contraception); Whalen v. Roe, 429 U.S. 589, 598–599 (1977) (noting privacy jurisprudence recognizes at least two types of interests, avoiding disclosure of personal information and independence in making certain kinds of important decisions); Lawrence v. Texas, 539 U.S. 538 (2003) (finding a Texas statute that criminalized sodomy intruded into the personal and private lives of individuals and violated the right to liberty under the Fourteenth Amendment). The Federal Trade Commission has committed to enforcing privacy promises even when the harm is not economic or physical or an unwanted intrusion, but merely unexpected disclosure of sensitive information about health or precise geolocation as well as less sensitive information such as purchase or employment history. FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY IN AN ERA OF RAPID CHANGE: RECOMMENDATIONS FOR BUSINESSES AND POLICYMAKERS, 13–14 (March, 2012), https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf.

81. *See* Daniel J. Solove, "*Nothing to Hide*" *and Other Misunderstandings of Privacy*, 44 SAN DIEGO L. REV. 745 (2007).

82. Notice and choice were two of the first Fair Information Principles developed by a government advisory committee addressing privacy concerns in the United States. *See* Robert Gellman, *Fair Information Practices: A Basic History* (Feb. 11, 2015), http://bobgellman.com/rg-docs/rg-FIPPShistory.pdf (documenting the history of the Fair Information Principles as well as their influence on U.S. laws).

Other approaches suggest privacy standards should reflect cultural norms,[83] and yet another view is that privacy might need to be imposed upon individuals by a paternalistic government.[84]

Disputes over the role of confidentiality also contribute to the instability of any comprehensive taxonomy. While privacy is generally considered to be about an individual's ability to avoid disclosure of his or her personal information, confidentiality is often used to describe a state of limited disclosure of that same information, perhaps to a person who has a duty to prevent further disclosure of conversations, such as an attorney providing legal advice. Some confidential relationships are recognized broadly throughout the law, while others are based on contractual principles or specific statutes that limit sharing of information.[85] Still other confidential relationships are supported by cultural, religious, or other social norms and have no enforcement mechanisms in the law.

---

83. The two-prong *Katz* test for violations of the Fourth Amendment includes both a subjective test for the defendant's expectation of privacy and an objective measure of the reasonable expectation of privacy. Katz v. United States, 389 U.S. at 347, 360–62 (Harlan, J., concurring). The reasonable expectation prong might well be an assessment of existing societal norms and realities. *See* Orin S. Kerr, *Four Models of Fourth Amendment Protection*, 60 STAN. L. REV. 503 (2007) (noting that societal understandings of privacy could be relevant to determining what constitutes Fourth Amendment reasonable expectations). In the online context, industry self-regulation for privacy is largely a measure of how much intrusion the market will tolerate without calling on Congress to formally regulate. *See* Omer Tene & J. Trevor Hughes, *The Promise and Shortcomings of Privacy Multistakeholder Policymaking: A Case Study*, 66 ME. L. REV. 437 (2014) (noting criticisms of industry codes of conduct and recommending structural supports to improve upon failed self-regulation).

84. *See* Anita L. Allen, *Coercing Privacy*, 40 WM. & MARY L. REV. 723, 755 (1999) ("Government will have to intervene in private lives for the sake of privacy and values associated with it.").

85. Evidentiary privileges against compelled disclosures support several confidential relationships including attorney-client, spousal, clergy-penitent, and physician-patient relationships. Edward J. Imwinkelried, THE NEW WIGMORE: A TREATISE ON EVIDENCE: EVIDENTIARY PRIVILEGES §3.2.4 (2014); *see also* Neil M. Richards & Daniel J. Solove, *Privacy's Other Path: Recovering the Law of Confidentiality*, 96 GEO. L.J. 123 (2007) (promoting the common law claim of breach of duty of confidentiality). Consumer enforceability of privacy policies has not been successful because of an unclear contractual status and difficulty in proving harm. A variety of scholarly proposals have emerged. *See, e.g.*, Woodrow Hartzog, *Reviving Implied Confidentiality*, 89 IND. L.J. 763 (2014); Joshua A. R. Fairfield, *"Do-Not-Track" as Contract*, 14 VAND. J. ENT. & TECH. L. 545 (2012). Much of privacy protection is conducted by the Federal Trade Commission through its authority to investigate and bring actions to address "unfair or deceptive trade practices," which have yielded some penalties for violations of privacy promises. *See* Daniel J. Solove & Woodrow Hartzog, *The FTC and the New Common Law of Privacy*, 114 COLUM. L. REV. 583 (2014).

The role of practical obscurity adds another element to the conceptual framing of privacy. Like confidentiality, practical obscurity creates expectations of limited disclosure based on practical barriers to sharing rather than on legal or social restrictions. Most of the debate about confidentiality and practical obscurity relates to the role that the law should play in supporting these social or practical norms.[86]

Context plays an important role in defining privacy, confidentiality, and practical obscurity, yet this is a difficult factor to encapsulate in a taxonomy of sensitive information. The contextual approach emphasizes that privacy risks vary based on the circumstances in which information is shared, including the relationships between the sharer and recipient as well as their expectations at the time of sharing. Time is also a contextual factor that can have an impact on both the harms and benefits that attach to the disclosure of sensitive information. Some approaches to privacy embrace the idea that the value of privacy increases over time compared with other interests,[87] and yet in other instances privacy interests are treated as decreasing with the passage of time.[88]

2. Charting the Piecemeal U.S. Approach to Privacy

Another challenge in creating a taxonomy of sensitive information is that different information types are treated as sensitive in different areas of the law. The piecemeal approach evident in U.S. privacy law is a function of the federal system, a history of legislating in response to startling events,[89] and the balancing of interests promoted by stakeholders. Many

86. *See* Nissenbaum, *supra* note 14, 155–56 (describing privacy norms as a function of many variables and suggesting that "protecting privacy will be a messy task").

87. The Court of Justice of the European Union held that Google must remove from its search results a link to a news article about a foreclosure that occurred more than a decade ago because the information was no longer timely. Case C-131/12, Google Spain SL v. Agencia Española de Protección de Datos (May 13, 2014), http://curia.europa.eu/juris/document/document.jsf?text=&docid=152065&doclang=EN.

88. *See* Douglas A. Kysar, *Kids & Cul-De-Sacs: Census 2000 and the Reproduction of Consumer Culture*, 87 CORNELL L. REV. 853, 870–75 (2001) (discussing the privacy concerns with census data collection despite assurances of confidentiality); HIPAA Privacy Rule, 45 C.F.R. § 164.502(f) (2015) (requiring covered entities to comply with the requirements of the HIPAA Privacy Rule for a period of fifty years following a decedent's death).

89. For example, Congress passed the Video Privacy Protection Act (VPPA), 18 U.S.C. § 2710 (2012), in 1998 following the disclosure of Robert Bork's video rental history during his nomination to the Supreme Court. *See* Neil M. Richards, *The Perils of Social Reading*, 101 GEO. L.J. 689, 694 (2013) (describing how "a horrified Congress quickly passed the VPPA," perhaps upon realizing that politicians' video rental records might otherwise be revealed as easily as Bork's); Andrea Peterson, *How a Failed Supreme Court Nomination Is Still Causing Headaches for Hulu and Netflix*, WASH. POST (Dec. 27,

U.S. privacy laws protect particular types of information within the context of a regulated sector, such as health care and banking, or within the context of limiting government power. The result is that particular information types may be protected in one sector although their privacy benefits can be outweighed by competing interests in another. Activity at the state level has also resulted in multiple approaches in areas such as data security breach notification requirements, and so the fragmentation is ongoing and pervades the U.S. legal system.

A related issue for the creation of a comprehensive taxonomy is that some sensitive information types are more clearly defined by law than others. Some laws are grounded in general principles like "unfair or deceptive trade practices"[90] or tautologies found in common law torts that provide civil remedies for the disclosure of "private facts."[91] These vague constitutional and tort protections for privacy contrast with health regulations such as the Health Insurance Portability and Accountability Act (HIPAA) that contains a list of seventeen sensitive information types that need to be redacted before regulated health entities can share personal health information.[92] Some privacy-related laws do not define sensitive information types at all and instead draw on influential policy statements,[93] industry standards,[94] and other areas of law.[95] Some U.S.

2013), http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/27/how-a-failed-supreme-court-bid-is-still-causing-headaches-for-hulu-and-netflix.

90. Federal Trade Commission (FTC) Act, 15 U.S.C. § 45 (2012). Much of privacy law now comes from the FTC's investigative and enforcement powers to bring or settle lawsuits when companies under their jurisdiction arguably fail to live up to their privacy promises. *See* Solove & Hartzog, *supra* note 85.

91. RESTATEMENT (SECOND) OF TORTS § 652D (1977) ("One who gives publicity to a matter concerning the private life of another is subject to liability to the other for invasion of his privacy, if the matter publicized is of a kind that (a) would be highly offensive to a reasonable person, and (b) is not of legitimate concern to the public.").

92. HIPAA Privacy Rule, 45 C.F.R. § 164.514(b)(2) (2015). The rule includes an eighteenth requirement, to remove "[a]ny other unique identifying number, characteristic or code." 45 C.F.R. § 164.514(b)(2)(i)(R) (2015).

93. *See* Gellman, *supra* note 82.

94. *See PCI SSC Data Security Standards Overview*, PCI SECURITY STANDARDS COUNSEL, https://www.pcisecuritystandards.org/security_standards/index.php (last visited Apr. 3, 2015).

95. *See* Nat'l Archives & Records Admin. v. Favish, 541 U.S. 157, 167–68 (2004) (finding that the Freedom of Information Act exemption for personal privacy extends to the familial concern in controlling the deceased's death images, in accordance with common law notions of privacy). Some state court electronic filing rules refer to state identity theft and breach notice laws, such as in North Carolina. *See N.C. eFiling Rules*, *supra* note 10. Rule 6.3 for e-filing in North Carolina defines "private information," as including sensitive, personal, or identifying information which must be excluded or

privacy laws are hybrids, with illustrative but non-exhaustive lists of protected information types.[96] Sometimes information is protected from some uses but not others, as in the Fair Credit Reporting Act, which limits the circumstances under which a consumer reporting agency can distribute consumer credit reports.[97]

Whatever benefits the sectoral approach brings, they are increasingly threatened by the ease with which information can be shared and aggregated. The increase in data brokers and the work of computer scientists and journalists highlight the leaky boundaries between separately regulated sectors and the potential for recreating previously redacted information by merging separate databases.[98] Information not restricted from disclosure in one context can obviate privacy protections in other parts of the dynamic information ecosystem. This development affects not just those individuals whose sensitive information is exposed through one sector but also those industry actors who invest in costly privacy and security approaches that prove to be ineffective. Public records in particular can spoil the privacy protections required in other areas because

partially described in court documents. The statutory basis for Rule 6.3 is N.C. Gen. Stat. § 132-1.10(d).

96. *See* Driver's Privacy Protection Act (DPPA), 18 U.S.C. §§ 2721–2725 (2012); Dahlstrom v. Sun-Times Media, 777 F.3d 937 (7th Cir. 2015) (holding that the DPPA's prohibition on disclosure of personal information in driving records did not raise heightened First Amendment scrutiny). The Freedom of Information Act lists exemptions, but refers to "personal privacy" somewhat unhelpfully. 5 U.S.C. § 552 (2012).

97. Fair Credit Reporting Act, 15 U.S.C. § 1681b (2012).

98. Journalists have been able to identify "anonymous" Internet users through records of their search history. Michael Barbaro & Tom Zeller, Jr., *A Face Is Exposed for AOL Searcher No. 4417749*, N.Y. TIMES (Aug. 9, 2006), http://query.nytimes.com/gst/fullpage.html?res=9E0CE3DD1F3FF93AA3575BC0A9609C8B63. Researchers have used "anonymous" Netflix viewing information released by the company to re-identify some of its customers. Steve Lohr, *Netflix Cancels Contest Plans and Settles Suit*, N.Y. TIMES (Mar. 12, 2010), http://bits.blogs.nytimes.com/2010/03/12/netflix-cancels-contest-plans-and-settles-suit. A graduate computer science student at MIT was able to re-identify the Governor of Massachusetts William Weld through presumptively anonymized state hospital records; the student recently reported forty percent re-identification capabilities in most contexts. Latanya Sweeney et al., *Identifying Participants in the Personal Genome Project by Name* (Apr. 29, 2013), http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2257732. The FTC's 2014 report on Data Brokers outlined a growing industry of data collectors and resellers who intermingle public records, information on the web, and proprietary data. FED. TRADE COMM'N, DATA BROKERS: A CALL FOR TRANSPARENCY AND ACCOUNTABILITY (2014), https://www.ftc.gov/system/files/documents/reports/data-brokers-call-transparency-accountability-report-federal-trade-commission-may-2014/140527databrokerreport.pdf.

they provide information that can be used for re-identifying individuals or for connecting a profile with information that carries privacy risks.

With these concerns in mind, we set out to create a sensitive information taxonomy that would allow for the identification of sensitive information in court records. We utilized a taxonomic approach for several reasons. First, we wanted to study the frequency of sensitive information in court records without first making a normative claim about what information types should be considered private. As a result, we cast the net widely and included a large number of information types, even those that appeared to have only a modicum of support in existing privacy law and scholarship. Second, we believe that the process we undertook to create our taxonomy will be valuable to other scholars and policymakers. In the sections that follow, we describe how we created our taxonomy and why we chose the information types that we did. Of course, not everyone will agree with our final list. Nevertheless, our taxonomy has proven to be a helpful guide in the assessment of the privacy risks that can arise from the public disclosure of court records, especially at a time when privacy laws have limited or unclear application to such records. Finally, even for those who disagree with our inclusions and exclusions, the instant taxonomy will serve as a useful starting point for the development of alternative taxonomies that scholars can apply to other information contexts.

B. CRITERIA FOR INCLUSION IN THIS STUDY

This project's taxonomy of sensitive information represents a broad list of information types that are protected by U.S. privacy law or that have been identified by scholars or others as information that should be protected from public disclosure.[99] To facilitate coding and analysis of the court records in our study, we grouped the various sensitive information types into the following thirteen categories:

1. Assets
2. Civil Proceedings
3. Computer Use
4. Criminal Proceedings
5. Education
6. Employment
7. Financial Information
8. Health

99. We conducted a survey of federal and state constitutional, tort, statutory, and regulatory law as well as federal and state court rules, European law, and legal scholarship.

9. Identity
10. Images
11. Intellectual Pursuits
12. Location
13. Sexual Activities

These are thematic categories that capture similarities in subject matter for the 140 sensitive information types that we searched for in the court records. Some information types logically fit in multiple categories, but we placed them in one category as a matter of simplicity for the coding and analysis of the records. There are, of course, other ways the individual information types can be categorized. We describe the makeup of each category in the sections that follow and provide a full listing of all of the coded information types in the Appendix.

After initial testing of the taxonomy, we decided to limit recording of sensitive information types to those occurrences that the coder could associate with an identified individual. In other words, we only coded for sensitive information that was associated with a person named in full or by last name within the brief or appendices of each document in the study. The identified individual did not have to be named on the same page where the sensitive information type occurred, but the association had to be clear to the coder from the information within the document.[100] For example, we would code for "Anne Klinefelter's Browning semi-automatic handgun" because it is apparent from the document that the gun is owned or possessed by Anne, but we would not code for "a Browning semi-automatic handgun was found in the street outside the grocery store" because the information is not associated with an identified individual.[101]

100. The only exception we made was for social security numbers because of their utility as a stand-in for personal identification. *See infra* note 139 and accompanying text.

101. Our requirement that sensitive information types needed to be associated with individuals named within the court briefs means that our findings do not include data that might directly support considerations of how discrete appearances of sensitive information (not associated with persons identified in court records) might be linked to individuals when court records are read in conjunction with outside sources. *See* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY, *supra* note 80, at 35–38 (suggesting that non-personally identifying information (PII) can be increasingly transformed into personally identifying information through re-identification); Christopher Wolf, *Technological Advances and Privacy Challenges*, *in* UNDERSTANDING DEVELOPMENTS IN CYBERSPACE LAW 23 (2014) (advising that traditional solutions focused on personally identifying information are undercut by big data practices and should be supplemented with techniques such as measuring risk of re-identification, and noting that publicly released data present greater risks for re-identification than data not publicly released); Paul M. Schwartz & Daniel J. Solove, *The PII Problem: Privacy and a*

We did not record occurrences of names alone, other than the names of minor children, absent some connection between the name and another piece of sensitive information (e.g., a person named as a juror, witness, or rape victim).[102] Where an individual was associated with another piece of sensitive information, we coded for whether the individual was an "adult," "minor," or "unknown." In privacy law, minors receive more protection than adults, so this status is itself a piece of sensitive information.[103] We did not code for information types associated with entities such as businesses, associations, and other groups.

What follows is an overview of the information types and categories in our taxonomy and some description of the sources that we used to create the taxonomy.

1. Assets

The "assets" category contains information relating to an identified person's possession or ownership of assets that might be considered sensitive, including financial assets, real estate, and vehicle identification numbers and license plate numbers. It also contains information indicating that an individual has a gun permit, filed a gun permit application, or possessed or owned a gun.

Scholars have noted that property ownership, including real estate ownership or rental status, constitutes information that can be used to identify individuals and that enables the creation of profiles used by data aggregators.[104] The Federal Trade Commission has noted that property records maintained by states are part of a growing data broker industry

New Concept of Personally Identifiable Information, 86 N.Y.U. L. REV. 1814, 1836–47 (2011) (noting that non-PII are no longer immutable categories due to the risk of re-identification and proposing an assessment using a continuum of identification risk).

102. We tested the coding of every appearance of a name early in the project, but so many names appeared in the briefs and appendices that it threatened to overwhelm our resources. We decided to leave this particular type of examination of court records to other researchers.

103. *See, e.g.*, Children's Online Privacy Protection Act (COPPA), 15 U.S.C. §§ 6501–06 (2012); REPORTERS COMM. FOR FREEDOM OF THE PRESS, PRIVATE EYES: CONFIDENTIALITY ISSUES AND ACCESS TO POLICE INVESTIGATION RECORDS 4–5 (2010), http://www.rcfp.org/rcfp/orders/docs/PRIVATEEYES.pdf (discussing protections for juvenile records); CAL. BUS. & PROF. CODE § 22581 (2015) (protecting minors online). Court electronic filing rules also generally allow for the replacement of the names of minors with their initials. *See supra* note 10.

104. Helen Nissenbaum, *Privacy in an Information Age: The Problem of Privacy in Public*, 17 LAW AND PHILOSOPHY 559, 561, 577 (1998) (listing information types traditionally treated as public and increasingly being used in the digital environment as identifying information in the organized surveillance of individuals).

that creates some risks of harm for consumers.[105] Vehicle identification numbers and license plate numbers are used in similar ways and are restricted under some state laws as well as the federal Driver's Privacy Protection Act.[106]

Information on gun ownership and possession is included because some states have passed legislation to protect the privacy of gun owners. For example, North Carolina exempts gun registration records from its public records law which would otherwise require public access to that information.[107] The Florida Firearm Owners Privacy Act limits a physician's ability to inquire about firearm access in patient interviews.[108] In addition, scholars raise the possibility that public disclosure of gun registration unconstitutionally burdens the right to bear arms.[109]

2. Civil Proceedings

The "civil proceedings" category is an organizing point for a number of types of information that relate to civil lawsuits and other non-criminal judicial proceedings. This category includes information relating to adoption, child support, civil commitment to a penal or mental facility, custody or guardianship proceedings, information indicating that an

105. FED. TRADE COMM'N, DATA BROKERS *supra* note 98, at 11–12 (reporting how state and local government records including property records are collected by data brokers either directly or indirectly); *see also* FED. TRADE COMM'N, PROTECTING CONSUMER PRIVACY, *supra* note 80, at 69 (recommending legislation to improve transparency in the data broker industry).

106. For example, North Carolina limits the sharing of vehicle identification numbers acquired through toll road administration. N.C. GEN. STAT. § 136-89.213 (2015). Regulations implementing the HIPAA Privacy Rule prohibit the sharing of vehicle identifiers and serial numbers, including license plate numbers, connected to personal health information. 45 C.F.R. § 164.514(b)(2)(i)(L) (2015). The federal Driver's Privacy Protection Act, 18 U.S.C. §§ 2721–2725 (2012), limits state departments of motor vehicles from sharing personal information except for specified purposes. The statute has been read to curtail distribution of "personal information," which is defined broadly. *See* Dahlstrom v. Sun-Times Media, 777 F.3d 937, 943 (7th Cir. 2015) ("[T]he DPPA's language appears broad: personal information means information that identifies an individual, . . . and there is no indication that Congress intended the enumerated list of examples to be exhaustive." (internal quotations omitted)).

107. N.C. GEN. STAT. §§ 14-415.17(c), 14-405(b), 14-406(a) (2015).

108. *See* Wollschlaeger v. Governor of Florida, 760 F.3d 1195 (11th Cir. 2014) (finding constitutional the effect of the Florida Firearm Owners Privacy Act on physicians' free speech rights); Act of April 26, 2011, 2011 Fla. Laws 112 (codified at FLA. STAT. §§ 381.026, 456.072, 790.338).

109. *See, e.g.*, Eugene Volokh, *Implementing the Right to Keep and Bear Arms for Self-Defense: An Analytical Framework and a Research Agenda*, 56 UCLA L. REV. 1443, 1548–49 (2009).

identified individual was the subject of a dependency or neglect proceeding, party to a divorce, juror name, and prior adverse civil judgments.

Information that falls within this category may be regarded as sensitive because individuals have no choice but to share these types of personal information in order to make use of government services or to remain law-abiding citizens.[110] The particular privacy implications of each of these information types are also advanced because of the special dignitary harms or, in the case of juror names, risk of retaliatory harms from public disclosure.[111] Information arising in civil proceedings that falls into a more specific category such as financial or health information is included in those more detailed categories below.

3. Computer Use

A number of information types that relate to an individual's use of computers or electronic information services comprise the "computer use" category: Instant Messenger or SMS identifier; IP address; Internet search history; Internet Service Provider (ISP) records including account number, billing information, or online access logs; computer password; Radio Frequency Identification (RFID); screen or user name for accessing a website or other online service; and Voice Over Internet (VOIP) username or number. These information types are culled from several different federal and state statutes as well as scholarship advocating protection for this kind of information.[112]

110. *See* Grayson Barber, *Personal Information in Government Records: Protecting the Public Interest in Privacy*, 25 ST. LOUIS U. PUB. L. REV 63, 71–72 (2006) (suggesting that court records "often contain information that is exquisitely personal"); Whalen v. Roe, 429 U.S. 589, 605–06 (1977) (holding that a state database on individuals with prescriptions for controlled substances did not violate a right to privacy because the data was not for public disclosure and was kept secure); Grayson Barber & Frank L. Corrado, *Public Access to Government Records and How Transparency Protects Privacy*, N.J. LAW., Oct. 2011, at 60 (protesting the government practice of selling personal information, and advocating transparency in order to generate citizen advocacy for greater privacy protection).

111. *See* Kenneth J. Melilli, *Disclosure of Juror Identities to the Press: Who Will Speak for the Jurors?*, 8 CARDOZO PUB. L. POL'Y & ETHICS J. 1 (2009) (advocating confidentiality of jury service to protect former jurors from harassment and physical threats); Kristin A. Henderson, *Lessons from Bankruptcy Court Public Records*, 23 LEGAL REFERENCE SERVS. Q. 55, 73, 76–77 (2004) (evaluating the particular sensitivity of information provided in bankruptcy proceedings).

112. *See, e.g.*, Children's Online Privacy Protection Act (COPPA), C.F.R. §§ 312.1–2; Omer Tene & Jules Polonetsky, *To Track or "Do Not Track": Advancing Transparency and Individual Control in Online Behavioral Advertising*, 13 MINN. J.L. SCI. & TECH.

4. Criminal Proceedings

Like the civil proceedings category, the "criminal proceedings" category serves as an organizing device for sensitive information related to the justice system. For this category, the information types are associated with law enforcement and criminal judicial proceedings, including information that identifies an individual as the subject of a criminal investigation, arrest, incarceration, conviction, sentence, or parole. The category also includes mug shots and pre-sentence investigation reports, sexual abuse allegations, child abuse allegations, and information concerning charges or convictions arising in juvenile proceedings. Additional information types are included for juror name, domestic violence victim name, rape victim name, and other crime victim name. The criminal proceedings category also includes cooperating defendant name, informant name, and witness name.

The information types we have listed within this category are widely regarded as sensitive.[113] For example, many scholars assert that the public disclosure of the names of crime victims and witnesses leads to the further victimization of those who have suffered from or witnessed criminal activity.[114] Others point to the stigma that attaches to individuals who have been subjected to criminal investigation, charge, or conviction.[115]

281, (2012); Chris J. Hoofnagle et al., *Behavioral Advertising: The Offer You Cannot Refuse*, 6 HARV. L. & POL'Y REV. 273 (2012).

113. *See, e.g.*, JAMES B. JACOBS, THE ETERNAL CRIMINAL RECORD 54–69 (2015) (describing the many types of criminal information in court records and criticizing their widespread availability); Sadiq Reza, *Privacy and the Criminal Arrestee or Suspect: In Search of a Right, In Need of a Rule*, 64 MD. L. REV. 755 (2005) (proposing increased privacy protections for the criminally-accused); Deanna K. Shullman & Mark R. Caramanica, *Mug Shots on Lockdown: Government and Citizen Backlash to "Exploitation" Websites Surges, Free Speech is the Casualty*, 30 COMM. LAW. 13 (2014) (surveying responses to businesses offering to take down mug shots for a fee and examining a split in federal circuit courts on the constitutionality of restrictions); Rebecca Hulse, *Privacy and Domestic Violence in Court*, 16 WM. & MARY J. WOMEN & L. 237 (2010) (examining the privacy rights of domestic violence victims in court and concluding that special protections should extend beyond family court contexts); Morrison, *supra* note 16 at 921 (highlighting the risks of harm from online court records in criminal cases and recommending the redaction of names of cooperating defendants and other informants while increasing transparency in the use of these law enforcement practices).

114. *See, e.g.*, Joel M. Schumm, *No Names, Please: The Virtual Victimization of Children, Crime Victims, the Mentally Ill, and Others in Appellate Court Opinions*, 42 GA. L. REV. 471, 486–93 (2008).

115. *See, e.g.*, Devah Pager, Bruce Western & Naomi Sugie, *Sequencing Disadvantage: Barriers to Employment Facing Young Black and White Men with Criminal Records*, 623 ANNALS AMER. ACAD. POL. & SOC. SCI. 195, 199 (2009), http://ann.sagepub.com/content/623/1/195 (finding that men with a felony drug conviction were fifty percent less

The combination of online and data broker exposure of often stale and incomplete arrest and conviction information is criticized as creating long-term barriers to fresh starts including negative impacts on "employment and housing prospects, parental rights, educational opportunities, freedom of movement, and just about every other aspect of daily life."[116] Broad exposure of and reliance on records of criminal activity is said to permanently mark ex-offenders as outlaws and restrict their ability to forge a path outside of crime, "a terrible outcome for society."[117] Information arising in juvenile delinquency proceedings has long been considered to be particularly sensitive because the disclosure of such information was "thought to hinder their rehabilitation by impairing their relations with the community [and] by stigmatizing them such that they view themselves as wrongdoers and act accordingly."[118]

As a result, a number of states have adopted or are considering broad sealing and expungement laws for various types of criminal information.[119] Although some of these laws will likely face significant constitutional challenges,[120] there is clearly a concerted effort by privacy and criminal justice advocates to limit the public disclosure of many types of criminal information.

5. Education

The "education" category encompasses five information types that relate to students at all levels of the education system: income eligibility for the National School Lunch program, the amount of financial aid awarded from federal or private sources, information indicating that a student was disciplined by his or her school, grades or other feedback from a school about a student's performance, and student identifiers.

Educational information is generally regarded as sensitive because it relates to a vulnerable class of individuals, often minors, who must share information with educational institutions, sometimes in a compulsory

likely than men without any record to receive a callback or be offered an entry-level job; black men with a record who applied were twice as likely as white men to be saddled with this "criminal record penalty").

116. *See, e.g.*, Jenny Roberts, *Expunging America's Rap Sheet in the Information Age*, 2 WIS. L. REV. 321, 327 (2015).

117. JACOBS, *supra* note 113, at 306.

118. Reza, *supra* note 113, at 785.

119. *See* Roberts, *supra* note 116, at 322.

120. *See supra* notes 25–30 and accompanying text.

education context. The information types included in this category are drawn from several federal statutory protections.[121]

6. Employment

Three information types are included in the "employment" category: information that an individual was disciplined by an employer, information describing an individual's military discharge, and performance evaluations of an employee. Information about the location where an individual works is included in the "location" category.

Employee privacy requirements vary by jurisdiction under statutory and common law.[122] In some states, performance evaluations are exempt from public disclosure.[123] The federal Freedom of Information Act includes an exemption for disclosure of contents of personnel files if that information would constitute a clearly unwarranted invasion of personal privacy,[124] and this exemption has been applied to performance appraisals.[125] Private employee privacy in performance appraisals varies, but in some states this information is protected by statute or by common law.[126]

121. The Family Educational Rights and Privacy Act protects student education records of institutions receiving federal funding. Family Educational Rights and Privacy Act, 20 U.S.C. § 1232(g) (2012); 34 C.F.R. § 99 (2015). The National School Lunch Act protects the confidentiality of the names of individual students who qualify for school lunch assistance. 42 U.S.C. § 1758(6) (2012).

122. *See* Pauline T. Kim, *Privacy Rights, Public Policy, and the Employment Relationship*, 57 OHIO ST. L.J. 671 (1996); *Access to Social Media Usernames and Passwords*, NAT'L CONFERENCE OF STATE LEGISLATURES, http://www.ncsl.org/research/telecommunications-and-information-technology/employer-access-to-social-media-passwords-2013.aspx (last visited Apr. 3, 2015) (listing recent legislation intended to strengthen workplace privacy regarding personal employee social media and other accounts); N.C. GEN. STAT. §§ 132-1.2, 160A-168 (2015).

123. States take different approaches to the accessibility of public employees' performance evaluations. *See* Roger A. Nowadsky, *A Comparative Analysis of Public Records Statutes*, 28 URB. LAW. 65, 86 (1996) (surveying states' laws and finding that in most states personnel files are presumptively private).

124. 5 U.S.C. § 552(b)(6) (2012).

125. *McLeod v. U.S. Coast Guard*, No. 96-5071, 1997 WL 150096 (D.C. Cir. 1997) (finding privacy interest in Coast Guard officer's evaluation report); Smith v. Dep't of Labor, 798 F. Supp. 2d 274, 284–85 (2011) (holding that disclosure of records containing performance appraisal information would constitute an unwarranted invasion of employee's personal privacy.)

126. *See* Laura B. Pincus and Clayton Trotter, *The Disparity Between Public and Private Sector Employee Privacy Protections: A Call For Legitimate Privacy Rights for Private Sector Workers*, 33 AM. BUS. L.J. 51, 54 (1995) (comparing states' approaches under statute and common law).

7. Financial Information

The "financial information" category contains a number of information types that relate to a person's financial condition and accounts. Separate types were specified for the account numbers associated with an individual's savings, checking, or other financial account; loan account numbers; credit card numbers; debit card numbers; and other types of financial accounts not already specified. Other information types in this category include information indicating that a person filed for bankruptcy or was adjudged to be bankrupt, that a person has been the subject of a foreclosure judgment, that a person owes a debt, or that a person has a lien on assets due to unpaid taxes. The ownership of physical financial assets such as stock certificates, cash, and coins is included in the "assets" category.

Tax returns are listed in this category as is information about compensation in the form of salary, wage, or other financial benefits, including stock options, court ordered payments, and other forms of compensation.[127] Insurance policy numbers, credit reports, and an individual's status as an identity theft victim are also considered sensitive information types under various laws.

Information that falls within this category is regarded as sensitive because it not only reveals details about a person's net worth, but it also may be useful in the commission of identity theft and consequential financial theft or credit harm. States have passed varying forms of restrictions on the sharing of financial and identifying information in order to reduce the risk of identity theft.[128] Other statutes such as the federal Fair Credit Reporting Act and the Gramm-Leach-Bliley Act provide protections for consumers seeking to restrict sharing of financial and related information.[129] In addition, debt and bankruptcy can result in

127. *See* Cynthia Blum, *The Flat Tax: A Panacea for Privacy Concerns?*, 54 AM. U. L. REV. 1241, 1262–81 (2005) (outlining a variety of harms that could result from inappropriate uses of tax information and recommending safeguards against disclosure and misuse).

128. The National Conference of State Legislatures maintains a chart of state data security breach notification laws intended to provide individuals an opportunity to monitor credit and financial accounts and to change passwords and credit card numbers to minimize the potential for identity-theft and related harms. *Security Breach Notification Laws*, NAT'L CONFERENCE OF STATE LEGISLATURES, http://www.ncsl.org/research/telecommunications-and-information-technology/security-breach-notification-laws.aspx (last updated Jan. 4, 2016).

129. Fair Credit Reporting Act, Pub. L. 91-508, Title VI, § 601, 84 Stat. 1114, 1128 (1970) (codified as amended at 15 U.S.C. §§ 1681–1681x (2012)); Gramm-Leach-Bliley

negative treatment even after an individual has regained financial stability.[130] The evolving law of court rules for electronic filing generally requires redaction of financial account numbers.[131]

8. Health

The "health" category includes information about abortion, cause of death, place of death, communicable diseases, dates of a hospital stay, disability status, drug or alcohol dependency, drug or alcohol treatment, HIV/AIDS status, and information relating to prescription medications. Paternity test information and pregnancy information are both in this category. Also included in this category are health plan beneficiary numbers, medical billing numbers, medical device identifiers or serial numbers, and medical record numbers. Other information in this category includes health diagnosis or treatment information not previously specified, genetic information, and medical conditions that are not the subject of diagnosis or treatment by a health care professional.

Information that falls within this category is regarded as highly sensitive because of the potential for discrimination based on perceptions of reduced capabilities or assumptions about unpopular causal behaviors.[132] Confidentiality of health information shared with a physician dates at least as far back as the Hippocratic Oath,[133] and is protected by many privilege laws and tort liability in some situations.[134]

The sources for information under the health category include the regulations authorized by HIPAA and its amendments,[135] privacy tort

Act, Pub. L. No. 106-102, § 527(4), 113 Stat. 1338, 1449 (1999) (codified as amended at 15 U.S.C. § 6827(4) (2012)).

130. Negative treatment of an individual based on debt that has been settled is sometimes considered an inappropriate response that should be prevented through restrictions in access to information about the debt. The right to be forgotten decision in Spain addressed this issue. See *supra* note 87.

131. *See supra* note 10.

132. *See* Charity Scott, *Is Too Much Privacy Bad for Your Health?* 17 GA. ST. U. L. REV. 481, 491–95 (2000) (reporting overwhelming public support for health privacy and articulating the harms of privacy violations as well as some benefits for certain exceptions).

133. Oath and Law of Hippocrates, circa 400 B.C.

134. Most states recognize an evidentiary privilege for physician-patient communications. *See* Edward J. Imwinkelried, THE NEW WIGMORE: EVIDENTIARY PRIVILEGES § 6.2.6 (2014); McCormick v. England, 494 S.E.2d 431, 437 (S.C. Ct. App. 1997) (reviewing the law of other states and joining the majority to recognize a tort for a "physician's breach of the duty to maintain the confidences of his or her patient in the absence of a compelling public interest or other justification for the disclosure").

135. The HIPAA Privacy Rule prevents the sharing of personal health information unless it is de-identified through the redaction of seventeen identifiers or through some

cases,[136] and federal constitutional law suggesting that information privacy may apply to some prescriptions.[137] The federal Genetic Information Nondiscrimination Act provides regulatory protection for genetic information.[138]

9. Identity

The "identity" category contains a number of information types that relate to an individual's physical and other characteristics that allow others to identify the individual. Like the other information types that we coded, information in this category must be associated with an identified individual. We did not code for the occurrence of full names or last names alone, even though we did use the appearance of associated names as the qualifying factor for almost all of the information types. We coded for Social Security Number (SSN) with or without names because SSNs are unique identifiers on their own.[139] We found that tracking names was simply too coder-intensive and provided inconsistent levels of information given that some names are common in the population and others are not. We did, however, code for the name of a minor child if it appeared in the

approved statistical approach. HIPAA Privacy Rule, 45 C.F.R. § 164.514(b)(2) (2015). The Driver's Privacy Protection Act prohibits states from selling or otherwise sharing drivers' "highly restricted personal information." 18 U.S.C. § 2721 (2012).

136. *See, e.g.*, Byrne v. Avery Ctr. for Obstetrics and Gynecology, 314 Conn. 433 (2014); Robert H. Thornburg, *Florida Privacy Law: Potential Application of Intentional Tort Principles and Florida's Constitutional Right of Privacy as Safeguards to Governmental and Private Dissemination of Private Information*, 4 FLA. COASTAL L.J. 137 (2003).

137. *See* Whalen v. Roe, 429 U.S. 589, 605 (1977) (holding that a New York database containing the names of individuals who were prescribed a controlled drug to treat depression did not burden a constitutional right to privacy because the statute had a rational basis and because the state adopted reasonable data security, but suggesting that a right to avoid wide disclosure of prescription information was at issue).

138. Genetic Information Nondiscrimination Act (GINA) of 2008, Pub. L. No. 110-233, 122 Stat. 881; 20 C.F.R. § 1635.1–12 (2015).

139. For a discussion of why we did not code for names alone, see *supra* note 102 and accompanying text. The perceived utility of social security numbers in facilitating identity theft and financial harm has inspired many states to pass legislation to restrict government collection of this information. *See, e.g.*, N.C. GEN. STAT. § 132-1.10(d) ("No person preparing or filing a document to be recorded or filed in the official records of the register of deeds, the Department of the Secretary of State, or of the courts may include any person's social security, employer taxpayer identification, drivers license, state identification, passport, checking account, savings account, credit card, or debit card number, or personal identification (PIN) code or passwords in that document, unless otherwise expressly required by law or court order, adopted by the State Registrar on records of vital events, or redacted.").

records because privacy laws provide special protection for children in a variety of contexts.[140]

The identity category includes driver's license number, email address, fax number, mother's maiden name, passport number, city and state of birth, professional certificate or license number, state identification number, and telephone number. Although we initially sought to code for gender,[141] we ultimately dropped this information type because it became unworkable. English language and naming conventions make gender cues so numerous that it was overwhelming the coding process.[142] We kept an information type for gender identity change, even though that designation is itself a contested and complicated issue.

This category also includes a number of information types associated with biological traits. Age, date of birth, date of death, barefoot print, fingerprint, gait, iris print or recognition, and voice print are included here, as are racial or ethnic origin. Some of this biological information is considered sensitive under a number of laws and is the subject of scholarship advocating increased privacy protection relating to the collection and use of biometric information.[143] Dates of birth and death

140. The Children's Online Privacy Protection Act provides special requirements for the collection of identifying information from children under the age of thirteen including first and last name. 16 C.F.R. §§ 312.1–3 (implementing 15 U.S.C. §§ 6501–6508). Federal Model Rule of Appellate Procedure 5.2(a)(3) provides that minor children may be identified with initials only. The North Carolina e-filing rules permits minors to be "identified by initials," *N.C. eFiling Rules*, *supra* note 10, at Rule 6.3; *see also* Schumm, *supra* note 114.

141. Federal and state statutes prohibit discrimination on the basis of gender in contexts such as employment and housing. *See, e.g.*, Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e (prohibiting employment discrimination based on race, color, national origin, sex, and religion); Fair Housing Amendments Act of 1988, 42 U.S.C. § 3604 (prohibiting housing rental or sale on the basis of certain traits including sex). Constitutional protections have also been recognized for gender nondiscrimination. *See* Craig v. Boren, 429 U.S. 190 (1976) (applying intermediate scrutiny to gender discrimination). While access to gender information is generally not restricted, requests for information relating to gender can create vulnerability for discrimination claims, so this information has been considered sensitive.

142. Pronouns alone triggered huge numbers of coding opportunities and created confusion about their meaning as neutral or gender-aware applications. In addition, coders might misread some names as conveying gender information.

143. The National Institute of Standards and Technology (NIST), an agency of the U.S. Department of Commerce, states, "Biometric technologies are used to establish or verify personal identity against previously enrolled individuals based upon recognition of a physiological or behavioral characteristic. Examples of biological characteristics include hand, finger, facial, and iris. Behavioral characteristics are traits that are learned or acquired, such as dynamic signature verification and keystroke dynamics." *Biometric Standards Program and Resource Center*, NIST, http://www.nist.gov/itl/csd/scm/

are restricted from disclosure under regulations implementing HIPAA. State identity-theft protection acts and court rules for electronic filing systems tend to require redaction of full birth dates.[144] Racial and ethnic origin define groups that are protected under the Fourteenth Amendment and by nondiscrimination statutes that generally place limitations on access to or use of this information.[145]

10. Images

The "images" category captures occurrence of photographs and video that contain a full-frontal view of an individual's face or features; show sexual organs of an undressed individual; show a person in a state of partial undress indicated to be taken without their consent; and photographs or videos depicting violence, abuse, or death of an individual. Video recordings that depict sexual acts are included in the "sexual activities" category.

These information types are drawn from protections provided through privacy torts, state statutes, and scholarship advocating additional protections for images, especially with the growing use of facial recognition.[146]

biometric-standards.cfm (last updated March 19, 2015); *see also* Margaret Hu, *Biometric ID Cybersurveillance*, 88 IND. L.J. 1475 (2013); Laura K. Donohue, *Technological Leap, Statutory Gap, and Constitutional Abyss: Remote Biometric Identification Comes of Age*, 97 MINN. L. REV. 407 (2012). Regulations implementing HIPAA restrict distribution of personal health information that contains biometric identifiers including finger and voice prints. 45 C.F.R. § 164.514(b)(2)(i)(P) (2015); *see also* N.C. GEN. STAT. § 14-113.20(b)(11) (2015).

144. *See* 45 C.F.R. § 164.514(b)(2)(i)(C); North Carolina Identity Theft Act, N.C. GEN. STAT. § 14-113.20.

145. *See* U.S. CONST. amend. XIV; Adarand Constructors, Inc. v. Pena, 515 U.S. 200, 201–02 (1995); Gratz v. Bollinger, 539 U.S. 244, 249–50 (2003); *see also* Title VII of the Civil Rights Act of 1964, 42 U.S.C. § 2000e (2012) (prohibiting employment discrimination on the basis of race, color, national origin, sex, and religion); Age Discrimination in Employment Act, 29 U.S.C. §§ 621–634 (2012) (prohibiting age discrimination against individuals over forty). States may also have nondiscrimination laws. While these statutes may not prohibit the gathering of information related to the protected traits, a strong defense against discrimination claims is that the sensitive information was not accessed.

146. *See, e.g.*, N.C. GEN. STAT. § 132-1.8. (protecting the confidentiality of photographs and video or audio recordings made pursuant to autopsy). Nonconsensual sharing of nude photographs or images of sexual activity are the subject of privacy torts claims and new statutes. *See* Danielle Keats Citron & Mary Anne Franks, *Criminalizing Revenge Porn*, 49 WAKE FOREST L. REV. 345 (2014); Clay Calvert & Justin Brown, *Video Voyeurism, Privacy, and the Internet: Exposing Peeping Toms in Cyberspace*, 18 CARDOZO ARTS & ENT. L.J. 469 (2000); Jeffrey R. Boles, *Documenting Death: Public Access to Government Death Records and Attendant Privacy Concerns*, 22 CORNELL J.L. &

11. Intellectual Pursuits

The "intellectual pursuits" category is a catch-all category that covers a range of information considered to be sensitive because it conveys information about the thoughts and views of individuals.[147] It includes cable television subscription records, cable television viewing history, video rental records, records of library use, and records of reading material purchased. Also in this category are the content of recorded conversations, political opinions, religious or philosophical beliefs, trade union membership, and voting information.[148] While activity on the Internet often reveals similar information,[149] we did not include computer-related identifiers in this category. Those identifiers were included under the computer use category.

A variety of statutes at the federal and state level address television viewing, video rental, and library use.[150] Records of books purchased may be protected as a constitutional right or through state legislation.[151] The

PUB. POL'Y 237 (2012) (examining the inconsistent treatment of death records including images in terms of access and privacy); Hu, *supra* note 143, at 1484–91 (explaining how face-recognition technologies and practices are growing while individuals are largely unaware of the privacy threats).

147. *See* Neil M. Richards, *Intellectual Privacy*, 87 TEX. L. REV. 387, 408 (2008) ("[W]e should understand intellectual privacy as a series of nested protections, with the most private area of our thoughts at the center, and gradually expanding outward to encompass our reading, our communications, and our expressive dealings with others.").

148. Federal and state wiretap laws offer protection against undisclosed recording of conversations in some contexts. *See, e.g.*, 18 U.S.C. §§ 2510–2522. Anonymity of speech is given some First Amendment protection, particularly in cases involving speech related to political activity and voting, religious freedom, trade union membership, and other associational activities. *See* NAACP v. Alabama *ex rel.* Patterson, 357 U.S. 449, 462–63 (1958) (holding that compelled disclosure of names of members would burden the right to freedom of association).

149. *See* Julie E. Cohen, *A Right to Read Anonymously: A Closer Look at "Copyright Management" in Cyberspace*, 28 CONN. L. REV. 981, 981–82 (1996); Richards, *supra* note 147, at 388–89.

150. The Video Privacy Protection Act, 18 U.S.C. § 2710 (2006), restricts video rental companies from sharing individuals' viewing habits. The Cable Communications Policy Act, 47 U.S.C. § 551(a)–(h) (2006), restricts disclosure of personally identifiable cable viewing records by cable television companies. Forty-eight states and the District of Columbia have statutes protecting some level of confidentiality of library use, and the remaining states, Hawaii and Kentucky, have Attorney General Opinions stating that the law of the state extends similar protection. The American Library Association maintains links to these state library confidentiality laws. *State Privacy Laws Regarding Library Records*, AM. LIBRARY ASS'N, http://www.ala.org/advocacy/privacyconfidentiality/privacy/stateprivacy (last visited Feb. 2, 2016).

151. Bookstores have asserted the confidentiality of books purchased. *See* Tattered Cover, Inc. v. City of Thornton, 44 P.3d 1044, 1053 (Colo. 2002) (finding that the Colorado constitution provided a higher level of protection than the federal constitution

confidentiality of associations that reveal unpopular political beliefs has been recognized as protected under the U.S. Constitution,[152] and scholars have advocated for recognition of other intellectual privacy protections.[153]

12. Location

Information in the "location" category includes geolocation information, home address, school address, and work address. Another information type in this category is full zip code with four or more digits that are associated with an identified individual.

At present, geolocation information is considered sensitive information requiring limits on disclosure under the Federal Trade Commission Act's Section 5 "unfair or deceptive trade practices" protections.[154] Geolocation information is considered sensitive when it is collected from a child using the Internet, and collection of this information is restricted under the Children's Online Privacy Protection Act.[155] Travel patterns evident in geolocation information are arguably sufficiently sensitive to merit Fourth Amendment recognition.[156]

---

when First Amendment and Fourth Amendment interests intersect in the case of law enforcement seeking book purchase records). California extends protection for e-reader privacy in the Reader Privacy Act, CAL. CIV. CODE § 1798.90 (2011).

152. *NAACP*, 357 U.S. at 463.

153. *See* Anita L. Allen, *Associational Privacy and the First Amendment: NAACP v. Alabama, Privacy, and Data Protection*, 1 ALA. C.R. & C.L. L. REV. 1 (2011) (reviewing the growth of protections for associational privacy, decisional privacy, and anonymity after *NAACP v. Alabama*); Neil Richards, INTELLECTUAL PRIVACY 5, 161 (2015) (examining law, policy, and practical approaches to "safeguard the processes of intellectual explorations and belief formation" and advocating for recognition that "*intellectual records are sensitive records* that demand higher protection than other kinds of data"). Additionally, the European Union provides protection for "personal data revealing . . . political opinions, religious or philosophical beliefs, [or] trade-union membership . . . ." Council Directive 95/46, art. 8, 1995 O.J. (L 281) 38 (EC).

154. *Prepared Statement of the Federal Trade Commission on S. 271, The Location Privacy Protection Act of 2014 Before the S. Comm. on the Judiciary Subcomm. for Privacy, Tech. and the Law*, 113th Cong. (2014) (statement of Jessica L. Rich, Dir. Bureau of Consumer Protection, Fed. Trade Comm'n), https://www.ftc.gov/system/files/documents/public_statements/313671/140604locationprivacyact.pdf; *FTC Casebook: Goldenshores Technologies, LLC & Eric M Geidl*, INT'L ASS'N PRIVACY PROF'LS, https://privacyassociation.org/resources/ftc-casebook/goldenshores-technologies-llc-erik-m-geidl (last visited Apr. 3, 2015) (providing case documents and analysis of FTC action regarding the deceptive use of geolocation data); *FTC Casebook: Aspen Way Enterprises*, INT'L ASS'N PRIVACY PROF'LS, https://privacyassociation.org/resources/ftc-casebook/aspen-way-enterprises (last visited Apr. 3, 2015) (discussing geolocation data).

155. *See* Children's Online Privacy Protection Act (COPPA), 15 U.S.C. §§ 6501–6506 (2012); 16 C.F.R. §§ 312.1–13 (2015).

156. *See* Hu, *supra* note 143, at 1481–82, 1500–03 (questioning the capacity of current Fourth Amendment jurisprudence to prevent growth in body and device tracking

13. Sexual Activities

The "sexual activities" category contains two information types: information about sexual activity and video or audio recordings of an identified individual engaged in a sexual act. Information and images relating to sexual activity are sometimes protected through privacy torts and through state statutes designed to address hidden cameras and non-consensual distribution.[157]

IV. STUDY DESIGN AND METHODOLOGY

To better understand the privacy interests that might be implicated by public access to court records, we selected a random sample of court records from a large corpus of North Carolina Supreme Court case files that are part of an ongoing digitization project by the UNC Law Library. We coded these documents in order to collect data about the frequency of appearance of sensitive information in the records, as well as other contextual information about the documents and the underlying cases. Once the coding was complete, we used statistical software to analyze the data we collected.

A. CORPUS OF COURT RECORDS UNDER STUDY

The UNC Law library has approximately 400 bound volumes of North Carolina Supreme Court case filings from 1928 to 2000. In 2013, the library embarked on an ambitious project to digitize some of these records and eventually make a version of them available and searchable online.[158] To date, the library has digitized 12,137 briefs and other filings from the North Carolina Supreme Court comprising 535,106 pages.[159] Each case

practices and detailing how geolocation tracking is expanding); Joel R. Reidenberg, *Privacy in Public*, 69 U. MIAMI L. REV. 141, 155–57 (2014) (proposing a new Fourth Amendment doctrine to address geolocational and other types of privacy).

157. Andrew J. McClurg, *Kiss and Tell: Protecting Intimate Relationship Privacy Through Implied Contracts of Confidentiality*, 74 U. CIN. L. REV. 887 (2006); MASS. GEN. LAWS ch. 272, § 105 (2015) (criminalizing the taking of "up-skirt" photos); N.J. STAT. ANN. § 2C:14-9 (West 2015) (criminalizing "revenge porn").

158. The North Carolina Supreme Court began providing electronic access to its filings in 2000, but does not provide electronic access to records from prior years. Copies of filings prior to 2000 were shared with several non-court libraries in the state, including the UNC Law Library.

159. For the period 1984–2000, the library has scanned and digitized the case fillings from 2255 cases heard by the North Carolina Supreme Court; this is approximately 85% of the cases in which the court issued a decision during this time period. For reasons unknown, the court did not send the library any case filings from approximately 15% of

file in this corpus contains at least one merits brief,[160] which may include an appendix. When a brief does have an appendix, it may contain court transcripts, witness testimony, and exhibits entered into evidence such as bank statements, medical records, psychological evaluations, and emails.

Our study used a stratified random sample by year of documents pulled from this corpus of records spanning the time period 1984 to 2000. These documents included briefs and petitions for discretionary review, along with their associated appendices. We did not review the "record on appeal," which is a separate filing containing, *inter alia*, copies of the case pleadings, jury instructions, transcripts, and other evidence filed in the lower courts.[161] In total, we analyzed 504 documents drawn from 466 cases.[162] One hundred and ninety-eight (39%) of these documents contained an appendix.

B. CODING AND ANALYSIS

We then performed content analysis on the documents in our sample.[163] This involved coding each document based on its content and case characteristics. The coding, which was performed by a team of eleven research assistants, captured information about each document (e.g., document type, length); information about the underlying case (e.g., date, case type); the type of sensitive information found in the record (e.g., social security number, HIV status); the general category of sensitive information (e.g., financial, health); and information about the location of

the cases the court heard and decided during this time period. These digitized records are being redacted in preparation for posting as searchable documents on the Internet.

160. In addition to a brief filed by the appellant, the case files also contain briefs by the appellee, reply briefs, and *amicus curiae* briefs. For cases that do not involve an appeal as of right to the North Carolina Supreme Court, the case file will also contain a petition for discretionary review.

161. *See* N.C. R. APP. P. 9. The UNC Law Library is not digitizing these records.

162. The number of documents exceeds the number of cases because some cases produced more than one document in our sample.

163. "Content analysis refers to the systematic reading and analysis of texts." Lee Petherbridge & R. Polk Wagner, *The Federal Circuit and Patentability: An Empirical Assessment of the Law of Obviousness*, 85 TEX. L. REV. 2051, 2070 (2007). In the context of legal scholarship, content analysis is typically performed on judicial opinions. *See* Mark A. Hall and Ronald F. Wright, *Systematic Content Analysis of Judicial Opinions*, 96 CALIF. L. REV. 63, 64–65 (2008). This methodology is also valuable in the analysis of court records because it allows us to empirically test scholars' intuitions about the content of court files. *See* David S. Ardia, *Free Speech Savior or Shield for Scoundrels: An Empirical Study of Intermediary Immunity Under Section 230 of the Communications Decency Act*, 43 LOY. L.A. L. REV. 373, 413 (2010) (noting that content analysis as a methodology allows scholars to move beyond anecdotes by generating objective, falsifiable, and reproducible data).

the sensitive information within the document (e.g., brief body, appendix).[164]

The selection and identification of sensitive information types was one of the central challenges of this study. As we described in Part III, there is no single, comprehensive list of private and sensitive information that we could utilize at the start of this project. Existing privacy laws, regulations, and customs have created a patchwork of inconsistent approaches and there is, as yet, no consensus among privacy scholars as to what information should be deemed private or sensitive in the context of court records.[165] Nevertheless, in order to make this study possible, we created a list of 140 sensitive information types based on a survey we conducted of existing legal authority and privacy scholarship and grouped those information types into thirteen categories.[166]

Once the documents were coded, we used STATA, a general-purpose statistical software package, to determine the frequency of appearance of each sensitive information type and to identify relationships, patterns, and correlations between different information types and other coded variables, including trends over time. A summary of our findings is included in Part V.

To check the reliability of the coding process we conducted two phases of testing.[167] First, we "pilot tested" a preliminary version of our coding form by having our coders review an identical set of five case documents.[168] We also reviewed those documents ourselves and compared the results of all coders. This resulted in minor alterations to the coding scheme and coder instructions. Second, we conducted a formal test of reliability at the conclusion of the coding process by selecting a random sample of fifty documents from the 504 documents in the study set.[169] We assigned each

164. The coders used Qualtrics, an online survey platform, to record their observations. The coding instrument and codebook are available on the authors' website. *See Media Law Resources*, UNC CTR. FOR MEDIA L. & POL'Y, http://medialaw.unc.edu/resources (last visited Feb. 1, 2016).

165. *See* Conley et al., *supra* note 13, at 775 (concluding that a complex body of rules, regulations, principles, and policies govern the creation of court records and access to them).

166. *See supra* Part III.

167. *See* Petherbridge & Wagner, *supra* note 163, at 2074 (noting that reliability testing is crucial because "the process of content analysis . . . is inherently subject to some level of subjectivity").

168. *See* LEE EPSTEIN & ANDREW D. MARTIN, AN INTRODUCTION TO EMPIRICAL LEGAL RESEARCH 101 (2014) (suggesting the use of a "pilot study" to pretest content coding schemas).

169. There is no bright-line standard dictating the sample size to be used when doing reliability testing. *See* Petherbridge & Wagner, *supra* note 163, at 2074, n.118 (stating

of the documents in this subset to a coder who had not previously coded the document. We then compared the results of the two codings in order to assess the degree of inter-coder reliability.[170]

V. RESULTS AND DISCUSSION

We begin in this part by presenting descriptive statistical information about the court records in our sample and the sensitive information they contain. We then examine the extent to which different types of sensitive information are related to various case and document characteristics.

A. DESCRIPTIVE STATISTICS

1. Sample Summary

The 504 court records that we reviewed contained a total of 24,156 pages, with a mean document length of 47.9 pages.[171] Table 1 presents a breakdown of the various document types that were in our sample, including the number of documents with an appendix and the length (in pages) for each document type.[172] As shown in Table 1, 198 (39%) of these documents included an appendix. Not surprisingly, documents that contained an appendix were substantially longer (mean length of seventy-one pages) than documents without an appendix (mean length of thirty-three pages).[173] We also found considerable variability among the

that researchers suggest that at least a ten-percent sample be used) (citing Stephen Lacy & Daniel Riffe, *Sampling Error and Selecting Intercoder Reliability Samples for Nominal Content Categories*, 73 JOURNALISM & MASS COMM. Q. 963, 969–73 (1996)).

170. The percentage rate of agreement and "Krippendorff's alpha," *see* KLAUS KRIPPENDORFF, CONTENT ANALYSIS: AN INTRODUCTION TO ITS METHODOLOGY 221–30 (2d ed. 2004), for each of the variables is listed on the coding form, which is available on the authors' website. *See Media Law Resources*, UNC CTR. FOR MEDIA L. & POL'Y, http://medialaw.unc.edu/resources (last visited Feb. 1, 2016).

171. The median document length was thirty-two pages. We report the median length in addition to the mean because document length was not normally distributed within the sample. The median document length is therefore a better measure of central tendency.

172. The document types were coded based on the document title on the first page of the brief or petition. According to the North Carolina Rules of Appellate Procedure, "The Title of the Document should reflect the position of the filing party both at the trial level and on the appeal, e.g., DEFENDANT-APPELLANT'S BRIEF, PLAINTIFF-APPELLEE'S BRIEF, or BRIEF FOR THE STATE." N.C. R. APP. P. app. E (1975). "Briefs for the State" are briefs filed by the State of North Carolina as either an appellant or appellee; the brief captions do not designate the specific role of the State.

173. These results suggest that there is a statistically significant difference between the distributions of page lengths for documents with and without an appendix ($z = -9.372$, $p = 0.0000$). We utilized the Wilcoxon-Mann-Whitney test for statistical

document types with regard to the inclusion of an appendix. Nearly all "Petitions for Discretionary Review" contained an appendix (98%)[174] while "Briefs for the State" were the least likely document type to include an appendix (16%).[175]

The most commonly occurring document type in the sample was briefs filed by the appellant, which constituted almost half of the documents (41%). The sample also included a number of non-party *amicus curiae* briefs (3%), which tended to be the longest documents in the sample. Although our sample did not include any reply briefs, we know from our review of the North Carolina Supreme Court's case files that a small number of reply briefs also exist in the population under study.[176]

significance because document length was not normally distributed within the sample. We return to the importance of the appendices in Part B.

174. The high proportion of petitions that included an appendix is likely due to the requirements in the North Carolina Rules of Appellate Procedure, which state that a petition for discretionary review "shall be accompanied by a copy of the opinion of the Court of Appeals when filed after determination by that court." N.C. R. APP. P. 15(c). We did not code for the type of documents attached as appendices, so we cannot state what proportion of the appendices included only a copy of the lower court's opinion.

175. The difference between these document types with regard to their inclusion of appendices is statistically significant (chi-square with five degrees of freedom = 97.9666, $p = 0.000$).

176. As with all random sampling approaches, there is a chance that our sample of documents did not capture the entire range of characteristics in the population of North Carolina Supreme Court records. It is likely, however, that any characteristics that are not in the sample appear very infrequently in the target population. Prior to 2009, the North Carolina Rules of Appellate Procedure did not permit the filing of a reply brief unless the court requested such a brief or certain special circumstances existed. *See* N.C. R. APP. P. 28(h) (1975).

Table 1: Frequency of document types in the sample, including the number of documents with an appendix [n, (%)] and length (in pages) for each document type.

			Page Length	
Document Type	***n***	**Appendix**	**Mean**	**Median**
Brief of Appellant	212	87 (41%)	54.4	34
Brief of Appellee	140	41 (29%)	37.8	30
Brief for the State	87	14 (16%)	46.0	30
Petition for Discretionary Review	46	45 (98%)	46.7	36
Brief of *Amicus Curiae*	16	9 (56%)	67.6	36
Other[177]	3	2 (67%)	30.7	26
All Document Types	504	198 (39%)	47.9	32

Our sample of documents came from cases decided by the North Carolina Supreme Court between 1984 and 2000, the years immediately preceding the introduction of electronic filing in North Carolina when additional rules regarding the redaction of sensitive information took effect.[178] Table 2 lists the number of documents in the sample by the type of case and the year in which the court issued its decision in the case. Because we selected a random stratified sample by year, the totals for each year are relatively constant, with a spike in the number of documents selected from cases decided in 1986 and a drop in documents from 2000. Although there was some variation in the proportions each year, nearly two-thirds of the documents in the sample came from civil cases (62%), slightly more than a third came from criminal cases (36%), and only a small proportion (1%) came from juvenile proceedings.[179]

177. The sample also included a motion to amend, guardian ad litem's brief, and brief by a cross-appellant.

178. *See supra* note 10 and citations therein.

179. Each state has special courts—typically called juvenile courts—that have jurisdiction over cases involving children under a specified age. *See* Juvenile Court, BLACK'S LAW DICTIONARY (10th ed. 2014). Juvenile court proceedings are civil as opposed to criminal. *Id.* In North Carolina, "[a] person who has not reached the person's eighteenth birthday and is not married, emancipated, or a member of the Armed Forces of the United States" is eligible for juvenile court if the case relates to abuse, neglect, or dependency. N.C. GEN. STAT. § 7B-101(14). Persons sixteen years and older who are charged with certain criminal violations or infractions are not eligible for juvenile court in North Carolina. *See* N.C. GEN. STAT. § 7B-1501(7) & (27). In our coding of the juvenile cases, we did not differentiate between delinquency cases and abuse, neglect, and dependency cases.

Table 2: Number of documents by case type and year of North Carolina Supreme Court's decision, including overall percentages for each case type.

	Case Type			
Year	**Civil**	**Criminal**	**Juvenile**	**Total**
1984	9	11	1	21
1985	16	12	0	28
1986	20	21	0	41
1987	23	11	1	35
1988	15	15	1	31
1989	11	12	0	23
1990	19	11	1	31
1991	24	10	0	34
1992	23	10	1	34
1993	20	9	0	29
1994	20	17	0	37
1995	23	9	0	32
1996	23	11	0	34
1997	20	10	1	31
1998	21	6	0	27
1999	18	5	1	24
2000	8	4	0	12
Total	313 (62%)	184 (36%)	7 (1%)	504

The cases themselves covered a wide variety of subject areas, ranging from appeals challenging death penalty sentences to workers' compensation determinations. To facilitate the coding of case subject areas, we adopted the twelve appellate case type designations created by the Court Statistics Project at the National Center for State Courts.[180] Not surprisingly, given that nearly two-thirds of the documents in our sample came from civil cases, the most commonly occurring appellate subject area

180. The appellate case types are: (1) Death Penalty; (2) Felony (non-Death Penalty); (3) Misdemeanor; (4) Criminal-Other; (5) Tort, Contract, and Real Property; (6) Probate; (7) Family; (8) Juvenile; (9) Civil-Other; (10) Workers' Compensation; (11) Revenue (Tax); and (12) Administrative Agency-Other. *See* COURT STATISTICS PROJECT, STATE COURT GUIDE TO STATISTICAL REPORTING 39–44 (2014), http://www.courtstatistics.org/~/media/Microsites/Files/CSP/State%20Court%20Guide%20to%20Statistical%20Reporting%20v%20211.pdf. The Court Statistics Project at the National Center for State Courts created these categories in order to provide a "standardized reporting framework for state court caseload statistics designed to promote intelligent comparisons among state courts." *Id.* at 1.

was "tort, contract, and real property," which constituted 32% of the documents in our sample. The next most common subject area was "felony (non-death penalty)," which arose in 26% of the documents, followed by appeals of administrative agency decisions, which appeared in nearly 9% of the documents in the sample. Documents from death penalty cases constituted 7% of the sample, yet they contained more than a quarter (28%) of the sensitive information we found.[181]

2. Sensitive Information Summary

Although a wide variety of sensitive information appears in the court records we sampled, it is not uniformly distributed throughout the records. Most of the documents contained relatively few incidences of sensitive information while a handful of documents contained a large number of pieces of sensitive information. Figure 1 presents a histogram of the frequency of sensitive information per document. It shows a pronounced rightward skew indicating that sensitive information is not "normally" distributed throughout the records.[182] In other words, the histogram is asymmetrical and does not have the classic bell shaped curve that would indicate that most documents fall within the middle of the range. Instead, the vast majority of documents contained fewer than forty pieces of sensitive information while only a few documents contained more than 400 pieces of sensitive information. At the far right of the graph we see that several documents contained more than 1,000 pieces of sensitive information. Overall, the records we reviewed contained an average of 113 appearances of sensitive information per document, with a median of thirty-six appearances of sensitive information.[183]

We saw considerable variation in the frequency of sensitive information among the different document and case types. Table 3 presents the median frequency of sensitive information by document type along with the location (brief body or appendix) where the information appeared.[184] Figure 2 presents similar information by case type.

181. We discuss the potential implications of this finding in Section VI.A.3.

182. A rightward skew is when the long tail is on the right side of the peak, which is also called a positive skew.

183. The standard deviation for the frequency of sensitive information coded per document is 209.07 and the interquartile range, covering the middle 50% of the observed frequencies, is 11–122.

184. The difference in the frequency of sensitive information between brief bodies and appendices is statistically significant (paired t-test with 104 degrees of freedom = -3.5484, p = 0.0006). We report the median frequency in Tables 3 and 4, rather than the mean, because the frequency of sensitive information was not normally distributed.

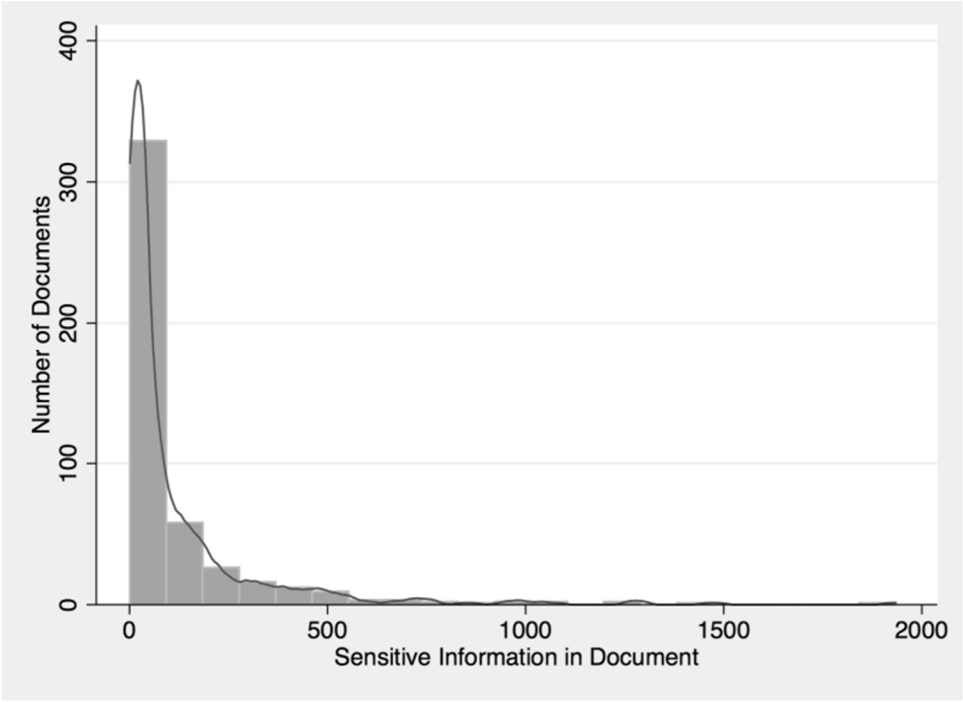

Figure 1: Histogram of frequency of sensitive information per document overlaid with kernel density plot.

Table 3: Median frequency of sensitive information coded per document, listed by document type and location within the document.

	Median Frequency of Sensitive Information		
Document Type	Brief Body	Appendix	Overall
Brief for the State	96.0	30.0	105.5
Brief of Appellant	39.0	24.0	42.0
Petition for Discretionary Review	14.5	17.0	22.0
Brief of Appellee	12.5	10.5	12.0
Brief of *Amicus Curiae*	8.0	215.5	10.5
Other	6.5	3.0	8.0
All Document Types	29.0	19.0	36.0

Figure 2 presents similar information by case type and reveals that criminal cases had substantially more sensitive information per document than either civil or juvenile cases. In fact, the median frequency of sensitive information in documents filed in criminal cases was approximately five times that of documents filed in either civil or juvenile cases. Figure 2 also reveals that in criminal and juvenile cases, sensitive information appeared

much more frequently in the brief body than in the appendix. In civil cases, sensitive information appeared with equal frequency in both appendices and briefs. We return to the role of appendices in Section V.B.

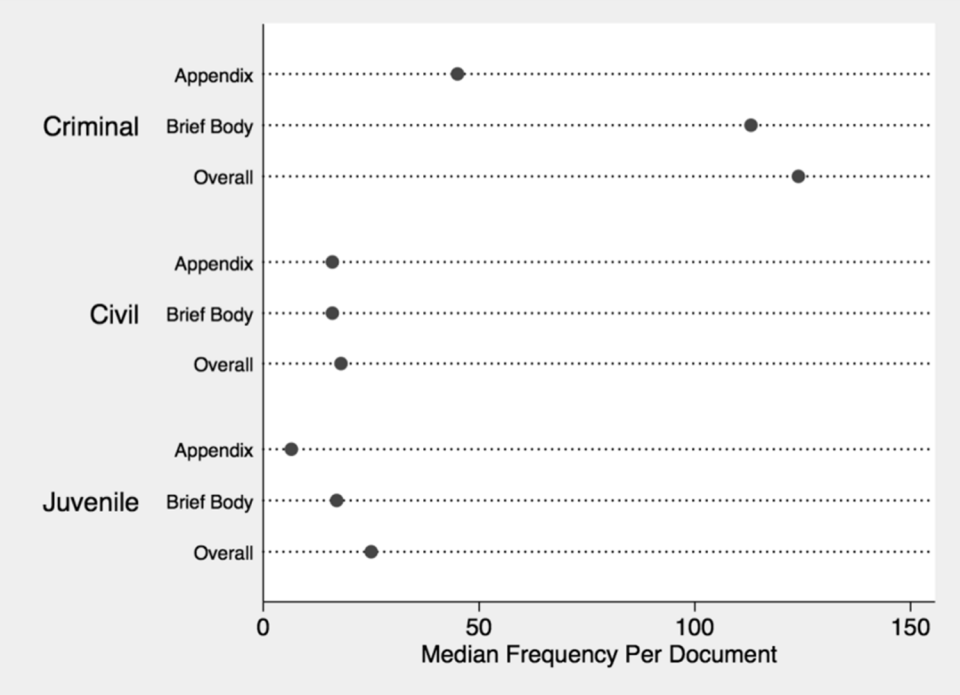

Figure 2: Dot plot of median frequency of sensitive information per document, by case type and location within the document.

As we noted in Part III, we grouped the specific sensitive information types into thirteen categories in order to facilitate comparisons between sensitive information types that shared similar characteristics. Table 4 reports the number of documents in the sample that contained sensitive information falling within each of these information categories. As Table 4 shows, information types in the "location" category appeared in more documents than any other category, appearing in 67% of the documents in the sample. Information in the "identity" and "criminal proceedings" categories also appeared in more than half of the documents, occurring in 66% and 56% of the documents, respectively. Overall, information in seven of the thirteen categories appeared in at least 20% of the documents.[185] Information in each of the remaining six categories appeared in fewer than 8% of the documents.

185. We break down the specific information types within these categories in Section V.B.

Table 4: Number of documents that contained sensitive information, listed by category of sensitive information, including the percentage of the total sample size (*n* = 504) and median number of times per document that information in that category appeared.

	Documents in Sample		Median
Information Category	***n***	**%**	**Per Doc**
Location	336	67%	4
Identity	331	66%	3
Criminal Proceedings	280	56%	54.5
Health	205	41%	5
Assets	175	35%	6
Financial Information	134	27%	4
Civil Proceedings	103	20%	4
No sensitive information	37	7%	-
Employment	33	7%	4
Sexual Activities	31	6%	6
Intellectual Pursuits	23	5%	4
Education	6	1%	3.5
Images	1	0%	17
Computer Use	0	0%	-

A few categories stand out in Table 4 because of their relative absence in the documents. Information about "sexual activities" appeared infrequently, as did information in the "intellectual pursuits" category, which includes religious beliefs, political opinions, and voting and reading records. Information about "education" was also mostly absent from the documents as were photos and videos captured by the "images" category. None of the documents contained any sensitive information in the "computer use" category (e.g., user names, passwords, and search history). Moreover, thirty-seven documents in the sample (7%) did not contain any of the sensitive information types that we coded for in this project.[186]

Table 4 also presents the median number of times per document that information in each category appeared. For most of the information categories, sensitive information appeared between three and six times per document. There are two outliers, however. Information in the "criminal proceedings" category appeared far more frequently in the documents than any other category (median appearance per document of 54.5), showing up approximately nine to eighteen times as often as information in the other categories. The "images" category was the other outlier, with a median of

186. A full list of the information types we coded is included in the Appendix.

seventeen appearances of sensitive information per document.[187] Although sensitive information in the "images" and "sexual activities" categories did not appear in very many documents, when they did appear, they generally did so with greater frequency than information in all of the other categories excluding "criminal proceedings."

B. ANALYSIS

It would be of little value to the debate over privacy and public access to simply add up the total number of times that various types of sensitive information appeared in court records. Indeed, we knew going into this study that court records are replete with sensitive information. Instead, in reporting and interpreting the results, we focus on the relative differences between document types, case types, and information categories rather than on the absolute numbers.

As noted in Part III and discussed more fully below, we purposefully coded a broad range of sensitive information types. Not all of the information that we identified presents the same privacy concerns. In the following sections we discuss the types of sensitive information that we found in the documents and examine the context in which they appeared.

1. Variations Within and Among Information Categories

Not surprisingly, the court records did not contain every category or type of sensitive information in equal measure. As Table 4 shows, information relating to location, identity, criminal proceedings, health, assets, finances, and civil proceedings appeared in many more documents than information that falls within the remaining six categories. We observed the same "top seven" categories of information when we calculated the total frequency of sensitive information throughout all of the records, but in a slightly different order.[188] For example, although the criminal proceedings category was only the third most frequently occurring information category on a per document basis (information types in this category appeared in 56% of the documents) it far exceeded every other category of information on the basis of total frequency of

187. This may be due to the fact that only one document in the sample contained information that fell within this category. A sub-sample consisting of only a single observation is too small to be statistically significant.

188. The top seven categories in terms of total appearance of sensitive information were criminal proceedings (n = 38,136), health (3,549), identity (3,217), assets (2,385), location (2,128), civil proceedings (1,428), and financial information (1,097).

appearance.[189] In other words, information related to criminal proceedings not only appeared in most court records, it also appeared more often in those records than any other category of sensitive information.

We might speculate that information related to criminal proceedings appears more frequently because criminal cases may be more common than civil cases. Documents from criminal cases, however, made up only 36% of the sample,[190] so the higher frequency of criminal information is not due to a larger number of criminal documents. Instead, criminal information was dispersed across all of the document types and case types. It is not just criminal cases that contain criminal information; this information appeared in a wide variety of contexts. We consider this further in the next section.

Turning to the individual information types in each of the most frequently occurring information categories, we saw a general pattern in the distribution of sensitive information. Figure 3 presents dot plots for the eight most frequently occurring information categories.[191] In each category, a few information types appeared far more often than the other information types in that category. This pattern was most evident for the financial information category, where information about an individual's compensation far outnumbered the other types of financial information in terms of frequency of appearance in court records.[192] This pattern was less pronounced for the assets and location categories, which had three and four types respectively of sensitive information that constituted more than 10% of their category's total. For the criminal proceedings and civil proceedings categories, the distribution was also more evenly spread; both of these categories had at least three information types that comprised 10% or more of their category's total.[193]

189. Information in the criminal proceedings category appeared 38,136 times in the sample. Information in the next highest category, health, had an overall frequency of appearance of 3,549. The substantially higher number of median appearances per document of information related to criminal proceedings as shown in Table 4 suggests this disparity as well.

190. *See* Table 2.

191. We present dot plots for the eight most frequently occurring categories, rather than just the top seven categories, to make full use of the space available in Figure 3. Note that the horizontal axes for these plots varied from a maximum frequency of 250 (for the "employment" category) to 15,000 (for the "criminal proceedings" category). The axes were presented in this way in order to allow clearer comparisons of frequency *within* each category.

192. There were surprisingly few incidences of bank account numbers (n = 9), credit card numbers (2), or other financial account numbers (4) in the sample. *See* Figure 3.

193. The criminal proceedings category dwarfed all other categories in terms of overall frequency of appearance. Unlike the other categories, six information types in the

As Figure 3 confirms, the criminal proceedings category far exceeded every other category of information on the basis of total frequency of sensitive information in the court records. Names of witnesses in criminal cases appeared more often than any other coded information type, followed somewhat distantly by the name of an individual who was the subject of a criminal investigation. The name of a victim of criminal activity other than rape (rape victim name is a separate information type) was the third most frequently occurring information type in the criminal proceedings category—and the third most frequently occurring information type overall. It is not until after the fifth most frequently occurring criminal information type, "conviction," that information in the other categories begin to place in the rankings of most frequently occurring sensitive information types.

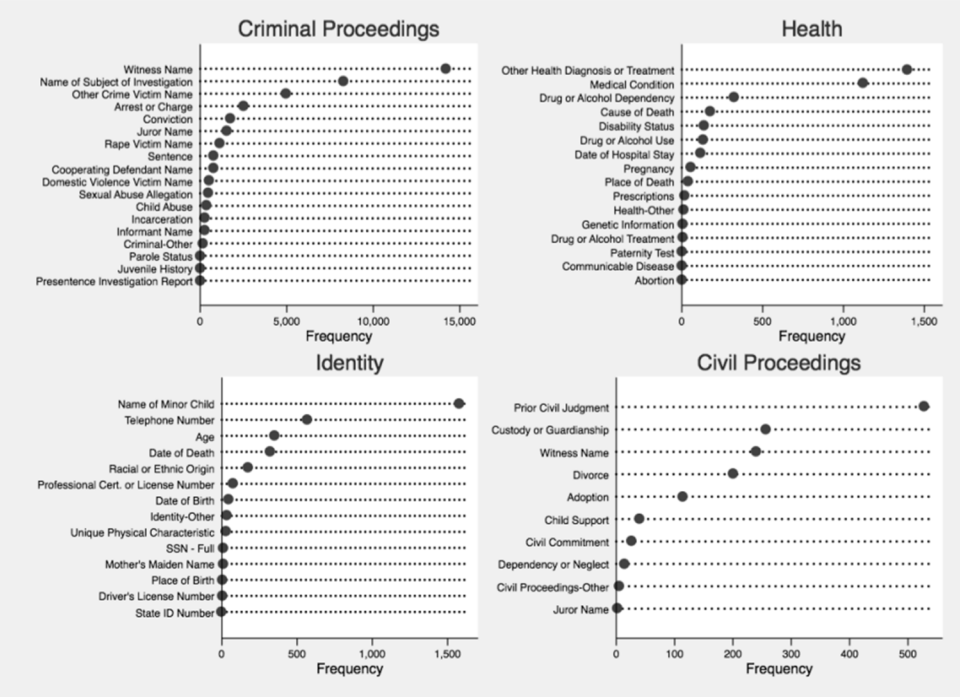

Figure 3A: Frequency of individual information types in the most commonly occurring categories of sensitive information.

criminal proceedings category appeared more than 1,000 times and three types of sensitive information appeared more than 4,900 times. *See* Figure 3.

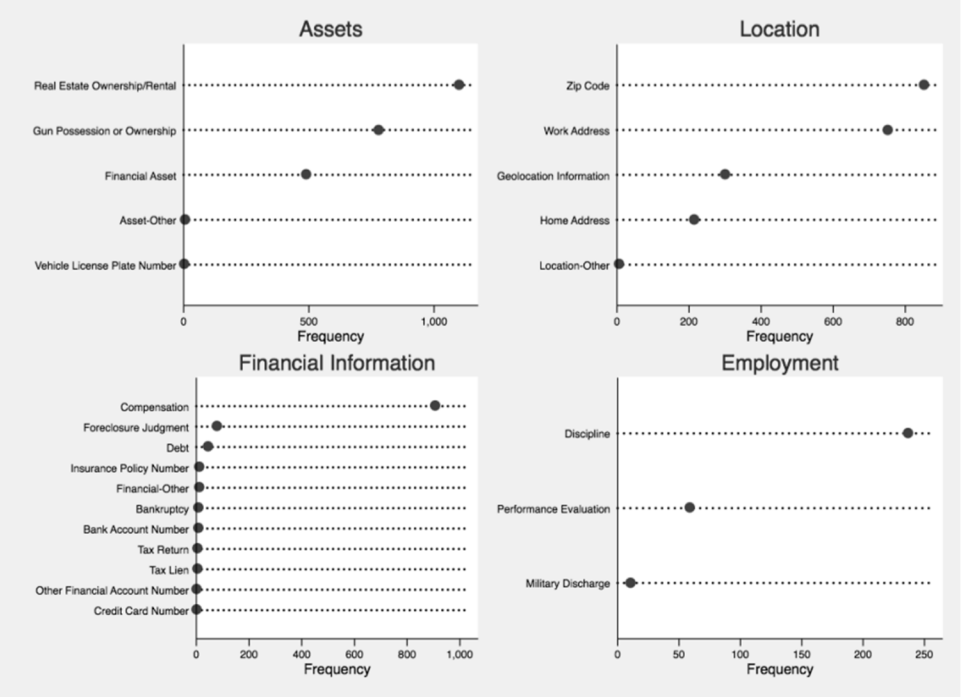

Figure 3B: Frequency of individual information types in the most commonly occurring categories of sensitive information.

The higher frequency of information related to criminal proceedings could be due to the fact that documents filed in criminal cases were, on average, longer than documents filed in other types of cases.[194] We would naturally expect longer documents to have more sensitive information. The data support this intuition, although the relationship between document length and frequency of sensitive information only partially explains the variations in the documents. Figure 4 presents a scatterplot of total sensitive information per document as a function of document length (in pages). The line through the scatterplot is the best-fitting linear regression line that provides an estimate of the relationship between the frequency of sensitive information in a document and the document's length.

194. Documents associated with criminal cases were on average 11.6 pages longer than civil cases: criminal cases had a mean [median] document length of 55.5 [36] pages compared to 43.9 [30] pages for documents filed in civil cases. Juvenile cases had on average the shortest documents, with a mean [median] of 32.3 [36] pages.

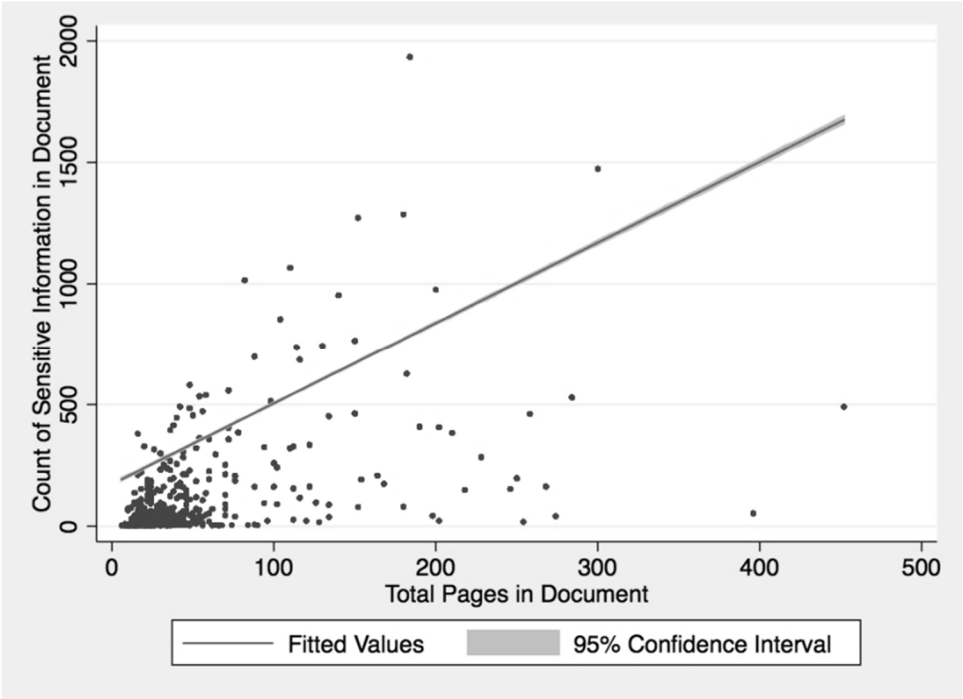

Figure 4: Scatterplot of frequency of sensitive information per document on document length (in pages) with linear regression line.

We draw several conclusions from Figure 4. First, the relationship between total frequency of sensitive information in a document and document length is positive (as documents get longer, we can expect to find more sensitive information). Second, the overall ratio is approximately 1:3 (for each additional page in length, we can expect to find approximately three more pieces of sensitive information).[195] Although as Figure 4 demonstrates, document length is an important indicator of the frequency of sensitive information in a court record, it accounts for only an estimated 33% of the variation in total frequency of sensitive information in the documents.[196] In other words, other independent variables, either alone or in combination, are likely to have a

195. Utilizing ordinary least squares, the regression model's coefficient for page length was 3.363195 (n = 52,998, std. err. = 0.0205702, R^2 = 0.3299, p-value = 0.000).

196. The linear regression model used in Figure 4 produces an estimate, known as the coefficient of determination (R^2), of the fit between the model's prediction of the number of appearances of sensitive information in a document as a function of page length and the actual frequency of sensitive information. For Figure 4, R^2 was 0.3299. This estimate tells us the percentage of the variance in the frequency of sensitive information explained by the model is 32.99%.

more substantial effect than page length on the frequency of sensitive information in a court record.[197]

Indeed, there are signs that other factors are at work when we look at scatterplots comparing the frequency of sensitive information as a function of document length across the three different case types. As Figure 5 shows, criminal cases had a higher density of sensitive information per page than either civil or juvenile cases. As page length increased, the number of pieces of sensitive information in criminal cases increased at a higher rate than it did in civil and juvenile cases. For criminal cases, the ratio between page length and frequency of sensitive information was roughly 1:4.[198] For civil cases, the ratio was approximately 1:1.[199]

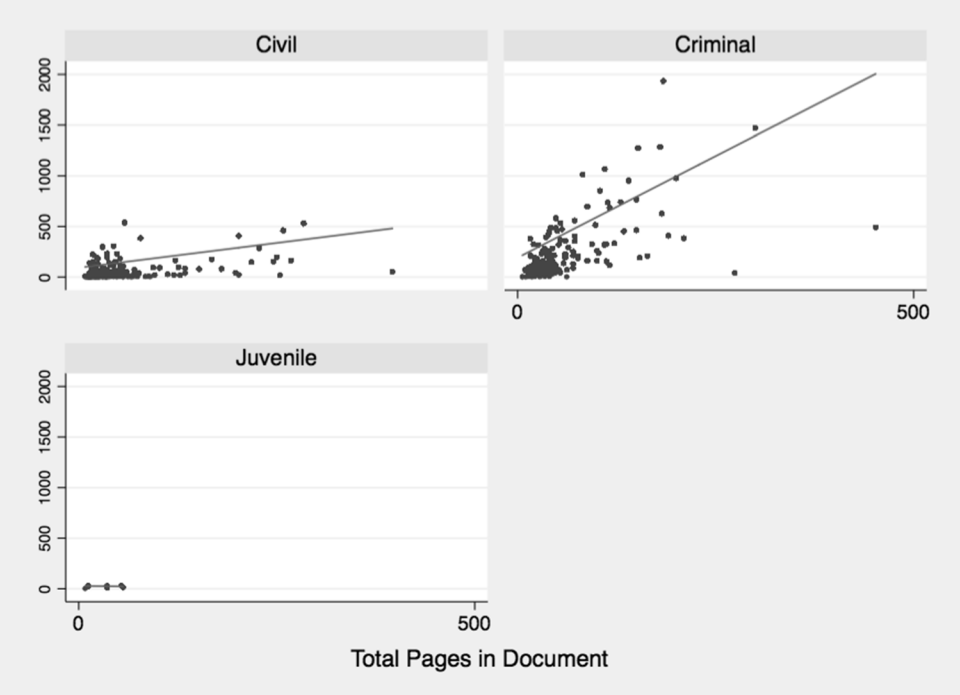

Figure 5: Scatterplot of frequency of sensitive information per document on document length (in pages) by case type with linear regression lines.

197. We report on the results of our multiple regression model in Part V.B.4.

198. Utilizing ordinary least squares, the regression model's coefficient for page length in criminal cases is 4.012137 (n = 40,440, std. err. = 0.0235338, R^2 = 0.4182, p-value = 0.000).

199. Utilizing ordinary least squares, the regression model's coefficient for page length in civil cases is 0.9802669 (n = 12,423, std. err. = 0.014221, R^2 = 0.2767, p-value = 0.000). There were too few documents from juvenile cases (n = 7) to draw any conclusions about the relationship between document length and frequency of sensitive information.

2. Contextual Variations

We turn now to the question of whether certain contextual factors influence—or at least are correlated with—the types of sensitive information found in court records. We coded for a number of case and document characteristics that might be linked to the appearance of sensitive information, including case type, case subject area, document type, subject of the information (adult or minor), location of the information (brief body or appendix), and year of case decision.[200] As the preceding discussion noted, we have already seen some variability in the types and extent of sensitive information associated with criminal cases, so we will start by analyzing the role that case type plays in the appearance of sensitive information.

a) Case Types

We know from Figures 2 and 5 that criminal cases have, on average, more sensitive information than civil and juvenile cases, but we cannot tell from those figures which types of sensitive information are more prevalent in criminal cases. Figure 6 presents the percentage of sensitive information in civil and criminal cases by category of sensitive information.[201] From Figure 6 we can discern some important differences about the extent of sensitive information in civil and criminal cases.

First, sensitive information is not uniformly distributed in all types of cases. The top bar in Figure 6 shows that overall, approximately 75% of the sensitive information we identified appeared in documents filed in criminal cases. Many of the information categories, however, deviated substantially from this 75/25 split.

In civil cases, we found a significantly higher proportion of sensitive information in the assets, civil proceedings, employment, financial, and location categories. In fact, sensitive information in the employment and financial categories appeared almost entirely in civil cases (92% and 94% of the time respectively). Sensitive information in the health, identity, and intellectual pursuits categories, on the other hand, appeared more frequently in documents associated with criminal cases, and information in the education and images categories appeared only in criminal cases. Only

200. Other contextual factors may also be relevant, but our focus here is on the case and document characteristics that courts themselves use in their filing systems.

201. Figure 6 does not include documents from juvenile cases because they were too few in number to warrant graphing.

information in the health and sexual activities categories appeared in roughly equal measure in both civil and criminal cases.

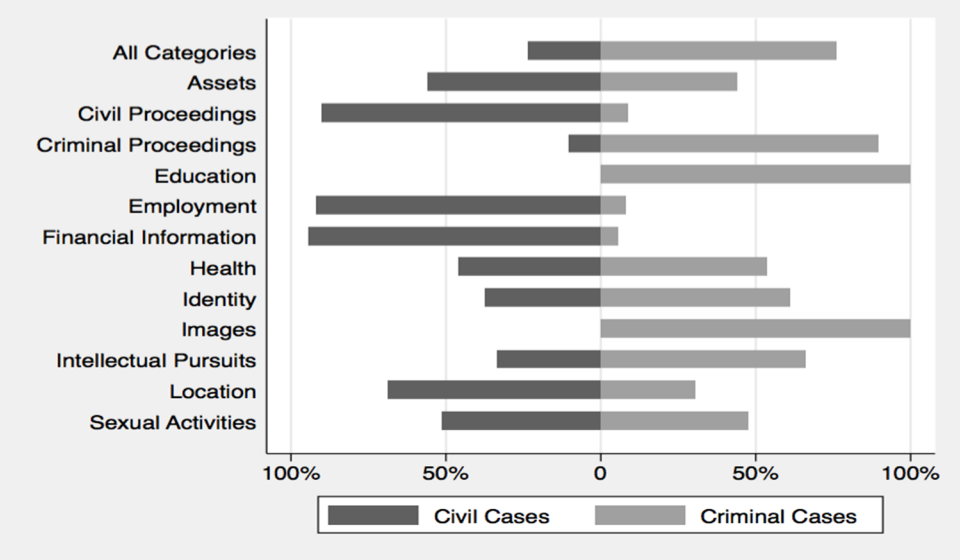

Figure 6: Horizontal bar graph showing percentage of sensitive information in civil and criminal cases by category of sensitive information.

Second, not only did criminal cases evidence more sensitive information than civil and juvenile cases, they also contained a greater variety of sensitive information. Criminal cases contained sensitive information from all of the categories we found in the documents,[202] whereas information from two categories, education and images, were absent from civil cases.[203] We also found that overall, criminal cases contained more types of sensitive information. Of the 140 sensitive information types we coded for in the records, ninety-five distinct types actually appeared in the documents. Although some types appeared exclusively, or nearly so, in civil cases, documents filed in criminal cases contained a greater variety of sensitive information types. Table 5 lists the information types that appeared in documents associated with only one

202. We did not find any information in the documents that fell within the "computer use" category. *See supra* Part III.B.3 (describing the information types in this category).

203. Six of the thirteen information categories were absent from documents filed in juvenile cases.

case type at least 90% of the time.[204] As Table 5 shows, criminal cases had a greater variety of sensitive information types than civil cases.

Table 5: Information types that appeared in documents associated with only one case type at least 90% of the time.

Civil Cases	Criminal Cases
Adoption	Abortion
Bank Account Number	Arrest or Charge
Bankruptcy	Child Abuse
Cable Television Subscription Record	Communicable Disease
Child Support	Content of Recorded Conversations
Compensation	Conviction
Debt	Credit Card Number
Discipline	Drug or Alcohol Treatment
Driver's License Number	Full-Face Photograph
Foreclosure Judgment	Genetic Information
Insurance Policy Number	Gun Possession or Ownership
Paternity Test	Incarceration
Performance Evaluation	Informant Name
Political Opinion	Juror Name
Prior Civil Judgment	Juvenile Court History
Professional Cert. or License Number	Military Discharge
SSN - Full	Name of Subject of Investigation
State ID Number	Parole Status
Tax Lien	Photos or Videos of Violence, Abuse or Death
Tax Return	Presentence Investigation Report
Voting Record	Rape Victim Name
	Sentence
	Sexual Abuse Allegation
	Student Discipline
	Student Grades or Performance Evaluation
	Vehicle License Plate Number
	Video Rental Records

204. None of the sensitive information types we coded for appeared more than 10% of the time in juvenile cases. There were several information types, however, that were disproportionately common in juvenile cases: "Adoption," "Custody or Guardianship," "Date of Birth," "Juvenile Court History," "Name of Minor Child," "Pregnancy," and "Sex Life."

b) Adults and Minors

Most of the sensitive information we found was associated with adults. Overall, only 7% of the sensitive information we coded was associated with an identified minor,[205] and the difference between civil and criminal cases with regard to sensitive information associated with minors was modest (information about minors appeared 6.1% and 7.2% of the time respectively). The percentage of sensitive information about minors was significantly higher in juvenile cases, where 40% of the information that we coded was associated with an identified minor.

Although appeals from juvenile cases are now subject to additional privacy protections under the North Carolina Rules of Appellate Procedure,[206] these protections were not in place during the time period we studied and we found a considerable amount of sensitive information in juvenile cases that was associated with an identified minor. Documents in juvenile cases contained an average of 10.27 pieces of sensitive information connected to an identified minor, with the following information types being the most common: "Name of Minor Child,"[207] "Rape Victim Name," "Adoption," "Age," "Arrest or Charge," and "Other Health Diagnosis or Treatment."[208]

We also found a few interesting differences between the information categories with regard to minors across all of the case types. Relative to their baseline percentages, minors were less likely to be associated with

205. For our purposes, an identified minor was any individual under the age of eighteen years, regardless of whether he or she met the requirements for juvenile court jurisdiction under North Carolina law. *See supra* note 179 (describing the jurisdictional requirements for juvenile court in North Carolina).

206. In 2006, the North Carolina Rules of Appellate Procedure were amended to provide additional privacy protections for juveniles. The current rules state that "covered juveniles . . . shall be referenced only by the use of initials or pseudonyms in briefs, petitions, and all other filings, and shall be similarly redacted from all documents, exhibits, appendixes, or arguments submitted with such filings" and that a "juvenile's address and social security number shall be excluded from all filings, documents, exhibits, or arguments with the exception of sealed verbatim transcripts." N.C. R. APP. P. 3.1(b).

207. As we noted in Part III.B, we coded for "Name of Minor Child" as a specific information type; this information type made up nearly 3% of the total frequency of all sensitive information in the records.

208. We did not record whether the minor in question was a "covered juvenile" under North Carolina law, so we cannot state whether the information we found would violate North Carolina Rule of Appellate Procedure 3.1(b). We can state, however, that we found no appearances in juvenile cases of addresses or social security numbers associated with minors. As we note in Part VI.A.4, the number of documents in our sample from juvenile cases was relatively small ($n = 7$), so we recommend further research on the privacy risks associated with minors.

information related to criminal proceedings (3.7%) and more likely to be associated with information in the education (33.3%), health (10.5%), identity (12.9%), and sexual activities (16.0%) categories.

There also were intriguing variations in the relative proportions of some of the specific information types with regard to minors. Figure 7 shows the percentage of sensitive information associated with adults and minors for the seventeen information types that were identified with minors more than 10% of the time. As Figure 7 reveals, a number of information types were disproportionately associated with minors (i.e., their association with minors was substantially greater than would have been expected based on the overall frequency of sensitive information associated with minors). Information about communicable diseases and minor names, for example, were exclusively identified with minors, and seven of the seventeen information types appeared more than 30% of the time in relation to a minor.

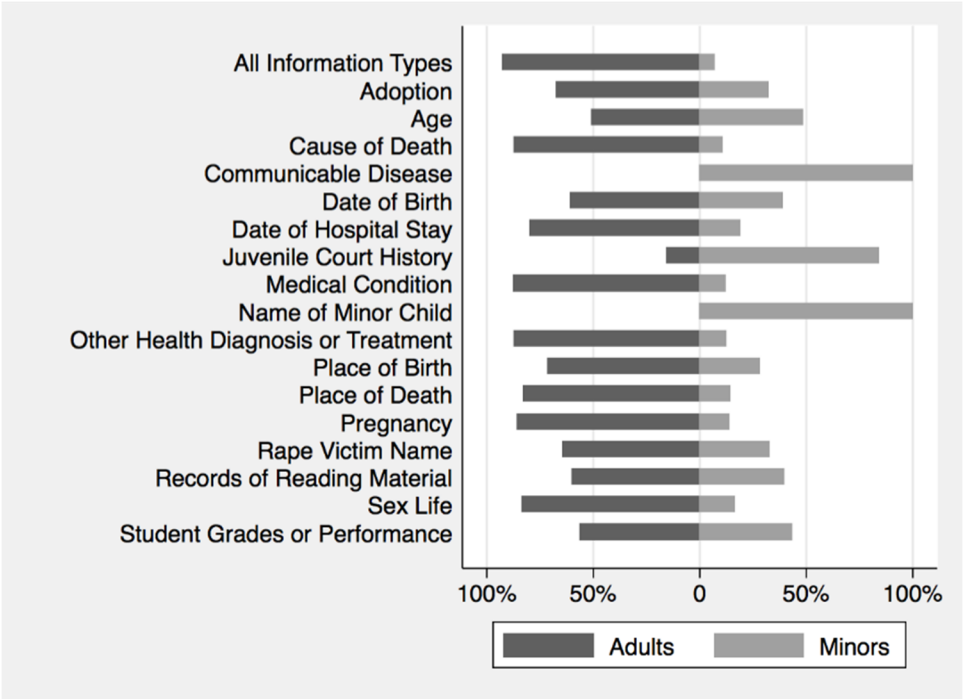

Figure 7: Horizontal bar graph showing percentage of sensitive information types associated with adults and minors. Only information types associated with minors more than 10% of the time are listed.

c) Appendices

Some scholars and archivists have suggested that appendices included in court records contain more sensitive information than legal briefs,[209] but our data did not bear this out. As we reported in Section V.A, we found that overall, brief bodies contained a higher frequency of sensitive information than appendices.[210] As the dot plot in Figure 2 showed, this disparity was particularly evident in criminal and juvenile cases; for civil cases, sensitive information appeared with equal frequency in both the appendices and brief bodies.[211]

Several information types, however, were more prevalent in appendices. Figure 8 lists the information types that appeared more than 30% of the time in appendices. Of the ninety-five information types we identified in the documents, sixteen appeared more than 30% of the time in an appendix, a marked deviation from the overall proportion of sensitive information in appendices (14%) as shown in the top bar in Figure 8. Moreover, seven information types appeared more than 50% of the time in an appendix, and three types appeared only in the appendices: "SSN - Full," "State ID Number," and "Video Rental Records."

As Figure 8 shows, only seven of the ninety-five information types that we identified in the records appeared more often in appendices. On the other hand, twenty-seven information types appeared exclusively in the brief bodies, including: "Abortion," "Adoption," "Bankruptcy," "Communicable Disease," "Credit Card Number," "Dependency or Neglect," "Drug and Alcohol Treatment," "Genetic Information," "Juvenile Court History," "Parole Status," "Paternity," "Student Grades or Performance," "Tax Lien," "Vehicle License Plate Number," and "Voting Record." Recall that nearly 40% of the documents contained an appendix.[212] Although there was substantial variability among the different

209. *See, e.g.*, Whiteman, *supra* note 17, at 470, 477 (describing the Northern Kentucky Law Library's decision to refrain from scanning appendices to briefs filed in the Kentucky Supreme Court); Hinderman, *supra* note 18, at 6 (noting that the Montana State Law Library pulled electronic court records it had already posted online to remove all exhibits and appendices before reposting the briefs).

210. *See supra* note 184 and accompanying text.

211. The median frequency of sensitive information in briefs and appendices filed in criminal cases was 113 and 41, respectively; for juvenile cases, the median frequency of sensitive information was 17 and 6.5, respectively. The median frequency of sensitive information in briefs and appendices filed in civil cases was 16.

212. *See supra* Table 1.

document types with regard to the inclusion of an appendix,[213] the variation between case types with regard to appendices was less pronounced and the differences were not statistically significant.[214] Accordingly, we can conclude that it is not a dearth of documents with appendices in our sample that is suppressing the appearance of sensitive information in the appendices. We will return to the role of appendices in Part VI.

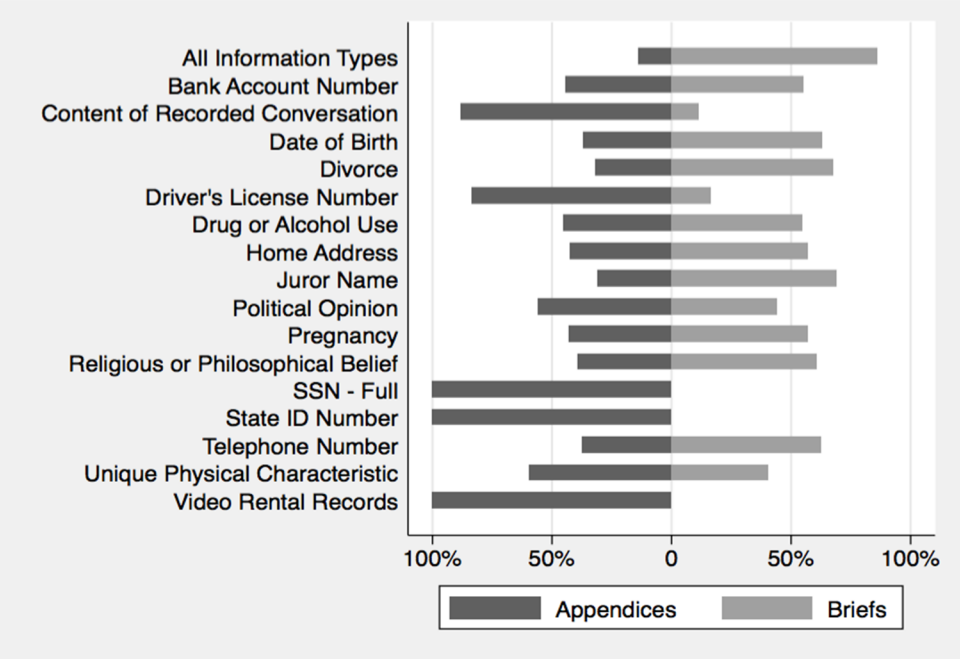

Figure 8: Horizontal bar graph showing percentage of sensitive information types in appendices and brief bodies. Only information types that appeared more than 30% of the time in appendices are listed.

3. Temporal Variations

Our final contextual factor is time. Following the approach of the North Carolina Supreme Court, which maintains its case files based on the year a case is decided by the court, we assigned the corresponding case year to each document in our sample. Figure 9 graphs the total number of

213. Nearly all petitions for discretionary review contained an appendix (98%) while briefs by the state were the least likely document type to include an appendix (16%). *See supra* Table 1.

214. Overall, 40% of civil cases included an appendix and 35% of criminal cases included an appendix. *See supra* Table 1. As noted in the text, the difference between the various case types with regard to the inclusion of appendices was not statistically significant (chi-square with 2 degrees of freedom = 2.1090, $p = 0.348$).

documents in our sample by year, as well as the number of documents that did not contain any of the sensitive information types we coded for. As Figure 9 shows, there was considerable year-to-year variation in both of these measurements.

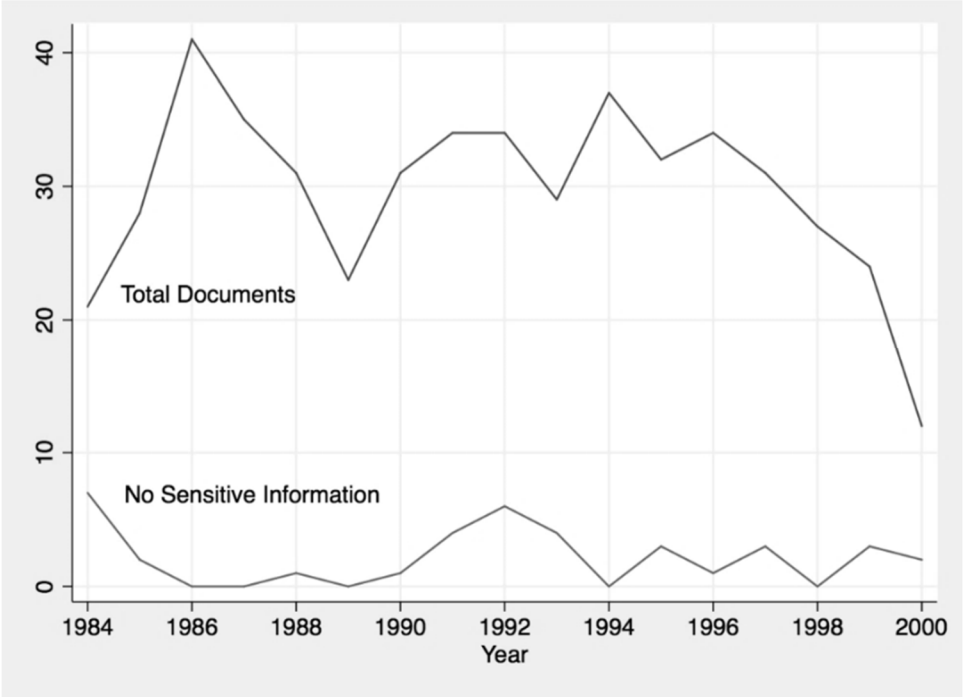

Figure 9: Total number of documents in sample and number of documents without sensitive information, by year.

We also found substantial variation in the total frequency of sensitive information per year. Figure 10 presents a two-way area graph of the total appearance of sensitive information by year as well as for the four most frequently occurring information categories: criminal proceedings, health, identity, and location. As Table 10 indicates, the total amount of sensitive information ranged from a low of 1,068 in 1984 to high of 6,052 in 1994, with an average of 3,117.5 pieces of sensitive information per year during the seventeen-year period under study.[215]

From Figure 10 we can see that in addition to variation in the overall frequency of sensitive information per year, the individual categories of sensitive information also varied during this time period. Not surprisingly,

215. The median frequency of sensitive information per year is 2,922; standard deviation is 1,700.3; and the interquartile range, covering the middle 50% of the observed frequencies, is 1,617–4,727.

information related to criminal proceedings tracked the overall totals quite closely (this is not surprising because the vast majority of sensitive information each year was associated with the criminal proceedings category). The other categories showed some variability as well, but did not parallel as closely the timing of the changes in the overall total. For example, the health category varied from a high of 626 in 1987 to a low of forty-four in 1997; the identity category varied from a high of 458 in 1995 to a low of fifty-seven in 1998; and the location category varied from a high of 305 in 1985 to a low of two in 1992.

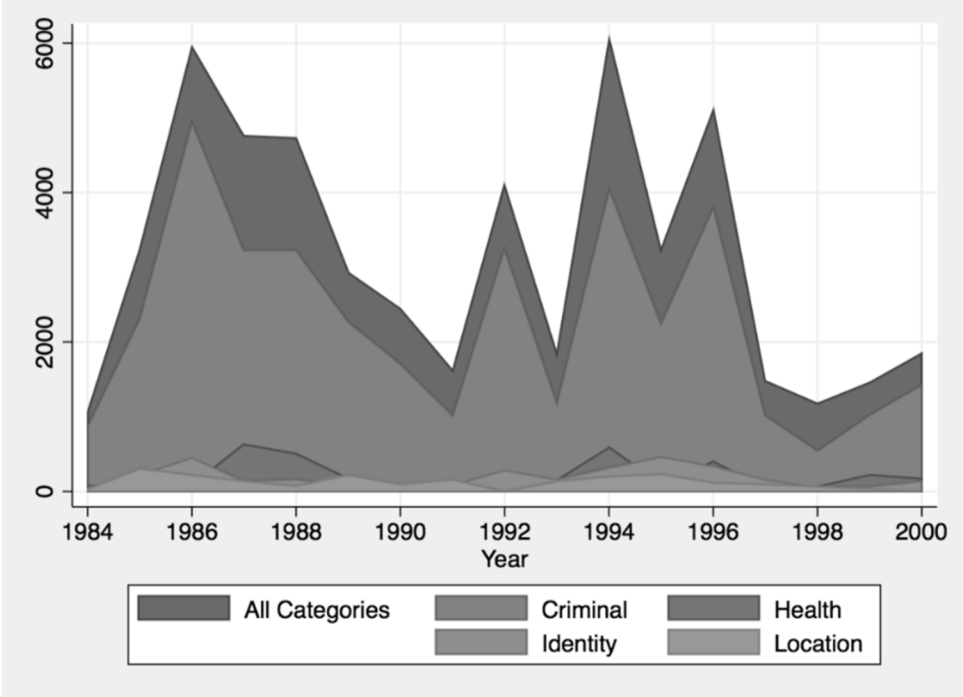

Figure 10: Total appearance of sensitive information by year of case decision including the 4 most frequently occurring information categories.

If we compare Figures 9 and 10, we might surmise that at least some of this variation is due to the fluctuations in the number of documents in our sample from each year. To remove this factor from our analysis, we calculated the frequency of appearance of the various categories on a per document basis. Figure 11 presents this measure of "sensitive information density" for the six most frequently occurring information categories.

Figure 11 suggests that there was no overarching trend in the appearance of sensitive information during the 1984 to 2000 time period. Instead, the numbers vary within a relatively constant range. It should be noted that the individual line graphs in Figure 11 do not all utilize the

same *y*-axis scale. The criminal proceedings category varied between twenty and 120 pieces of sensitive information per document whereas the location and civil proceedings categories varied between zero and ten. By varying the *y*-axis scale, Figure 11 allows us to compare the relative changes in sensitive information density *within* each category over time and leads to two important observations.

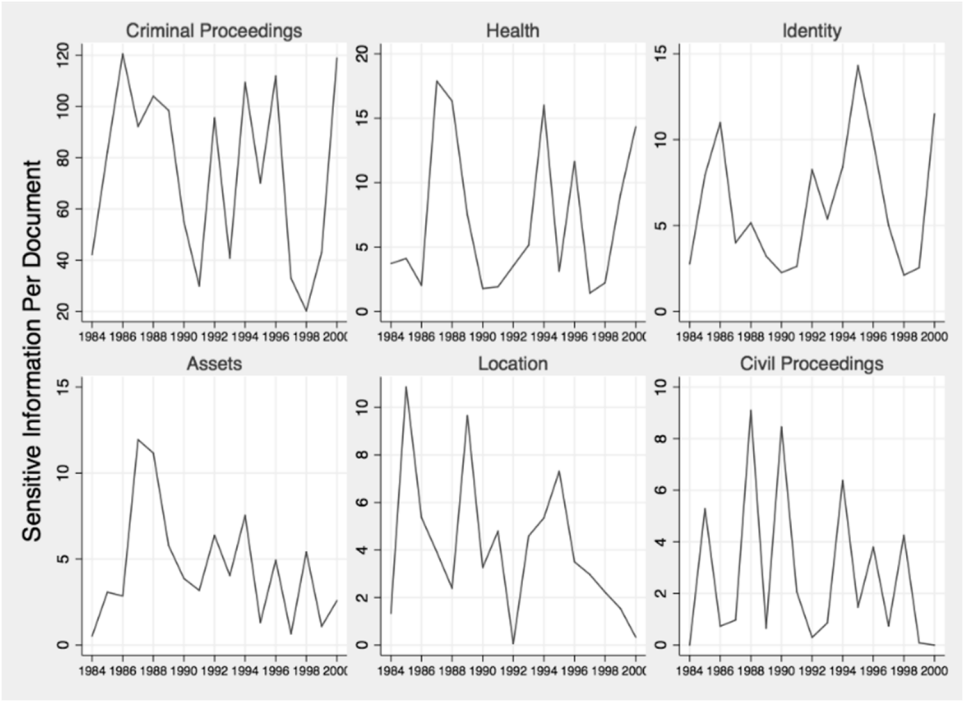

Figure 11: Sensitive information per document by year of case decision for the six most frequently occurring information categories.

First, there was considerable variability both within the categories and among the categories during this time period. None of the categories evidenced a consistent amount of sensitive information per document on a year-to-year basis. In some years there were sharp increases in the amount of sensitive information associated with these categories while in other years there were steep declines. For example, the amount of sensitive information per document in the health category spiked in 1987 and 1984 and fell in 1990 and 1997, whereas the identity category showed sharp increases in 1986 and 1995 and declines in 1990 and 1998. Moreover, the peaks and valleys evident in the individual graphs in Figure 11 do not align. When some categories were peaking, others were dipping.

Second, we do not see either a declining or rising trend for these categories. Although there is a pronounced increase in sensitive

information per document beginning in 1998 for the criminal proceedings, health, and identity categories, the location and civil proceedings categories declined during that same period, and overall the five-year moving averages for all of the categories show no discernable trend. From Figure 9 we can see that the upswing in sensitive information for the criminal, health, and identity categories in 1998 occurred during a period when the total number of documents was declining. Whether this increase in sensitive information continued after 2000 is beyond the scope of this study.

4. Regression Analysis

In this section we use multiple regression analysis to examine whether and to what extent certain document and case characteristics influence the amount of sensitive information in court records. As we previously noted, document length is a statistically significant predictor of the amount of sensitive information in a court record.[216] Recall that the linear regression line shown in the scatterplot of sensitive information per document in Figure 4 provided an estimate of the relationship between the frequency of sensitive information in a document and the document's length.[217] We now add other independent variables to our analysis in order to better predict the amount of sensitive information in court records.

Figure 12 presents a nomogram of the multiple regression coefficients for the analysis of the frequency of sensitive information per document (log transformed) for nine independent case and document variables.[218] It shows that six independent variables—criminal case type, appellant's brief, appellee's brief, petition for discretionary review, state's brief, and document length (in pages)—are statistically significant predictors of the total amount of sensitive information in a document because their 95%

216. *See supra* note 195 and accompanying text.

217. The best fitting linear regression line in Figure 4 predicted a relationship of approximately 1:3 between document length and frequency of sensitive information, with a coefficient of determination, R^2, of 0.3299. *See supra* note 195 and accompanying text.

218. We utilized a log transformation of the total amount of sensitive information per document because this variable is not normally distributed. Each parameter estimate in Figure 12 is represented by a dot along with the 95% confidence interval for the estimate depicted by a horizontal line. Parameters with narrower confidence intervals are estimated more precisely than those with wider confidence intervals. Only those parameters with a confidence interval that does not cross zero are statistically significant at a 95% level. Figure 12 does not include the intercept parameter, which has a coefficient of 0.6542635 [95% CI: -1.089448 to 2.397975].

confidence intervals do not cross zero.[219] In other words, holding all other variables in the model constant, the amount of sensitive information in a document can be predicted based on the document's length, the type of brief (other than an amicus brief), and whether it was filed in a criminal case. The other variables listed in Figure 12 may also influence the amount of sensitive information in a document, but the findings from the multiple regression model do not show that we can be sufficiently confident to assess what effect, if any, they have on the amount of sensitive information.

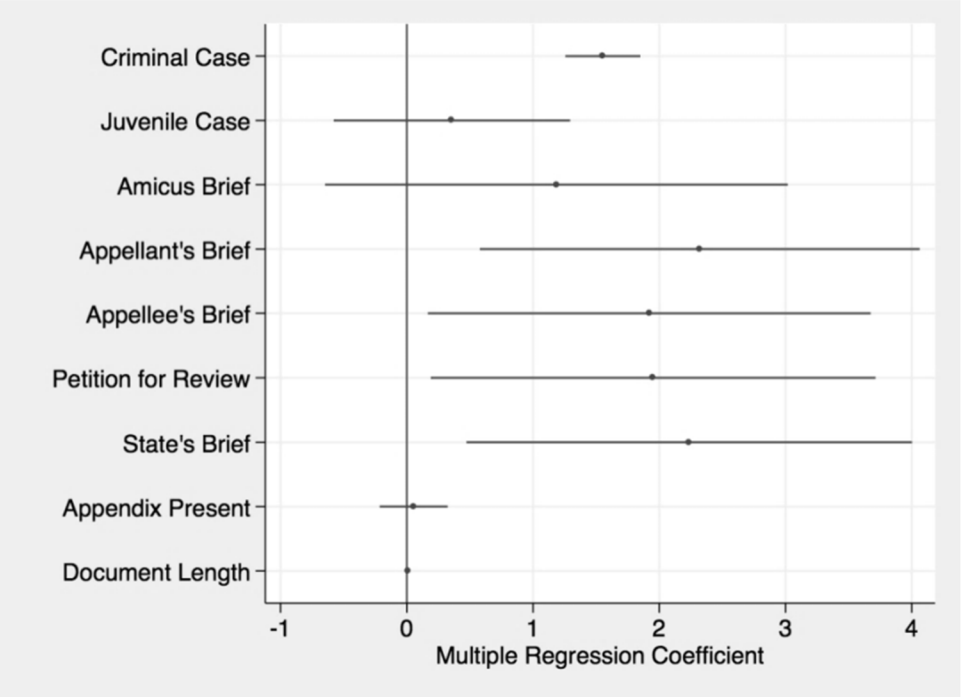

Figure 12: Nomogram of multiple regression parameters and 95% confidence intervals for the analysis of sensitive information per document (log transformed) listed by case and document variables.

What the multiple regression model tells us is that, other things being equal, criminal cases contain significantly more pieces of sensitive information per document than civil cases (coef. = 1.55, $p = 0.000$). In addition, appellee briefs (coef. = 1.92, $p = 0.032$), petitions for discretionary review (coef. = 1.95, $p = 0.030$), briefs by the state (coef. =

219. The model's other relevant statistics include $F(9,457) = 43.71$, $p < 0.0000$, $R^2 = 0.4626$. All six independent variables described in the text added statistically significantly to the prediction, $p < 0.05$. The full regression table is included in the Appendix.

2.23, $p = 0.013$), and appellant briefs (coef. = 2.32, $p = 0.009$) contain significantly more pieces of sensitive information than the other briefs category (reference group), with the relative impact of the various brief types increasing as their coefficients increase. It also tells us that the number of pages in a document is another significant predictor of sensitive information. The more pages in a document, the more pieces of sensitive information appear in the document (coef. = 0.01, $p = 0.000$). As for the inclusion of an appendix, the model shows that this does not predict the overall frequency of sensitive information in the document.

This final point about the lack of impact of appendices may strike some readers as surprising, but given our earlier finding that brief bodies contain a higher median frequency of sensitive information—as well as a wider variety of sensitive information types—than appendices,[220] it is not unexpected.[221] We will return to the implications of these findings in Part VI.

VI. IMPLICATIONS FOR ACCESS POLICIES AND PRACTICES

As we noted in the introduction, courts and libraries are moving quickly to digitize court records and make them available online. The results of this study will, we hope, inform these efforts by providing much-needed detail about the extent and context of sensitive information in these important public records. In this part we discuss how our study can aid the ongoing debate about privacy and court records and how our results can help to identify and remedy potential implementation challenges if courts or archivists decide to carry out privacy management protocols.

It is not our goal in this Article to tell courts or archivists what information, if any, should be redacted or what documents should be withheld from online access or otherwise managed for privacy protection. These largely normative questions must be answered based on a careful balancing of the competing public access and privacy interests. Moreover, the privacy interests cannot be evaluated based solely on the presence or absence of specific types of sensitive information in individual court records. Other factors, including what one can learn or infer about individuals from other sources, as well as the value to society of the

220. *See supra* note 184 and accompanying text.

221. Another possible explanation is that the inclusion of an appendix is highly correlated with page length.

information in question, must also be taken into account. The data we present will be invaluable in doing this evaluation, but our findings should not be read to dictate one approach or another.

As discussed in Part IV, our sample of court documents came from a large corpus of briefs and other filings submitted to the North Carolina Supreme Court spanning the time period 1984 to 2000.[222] Although our results are specific to this population of court records, we believe that the data we collected shed light on court records held by other state appellate courts, particularly courts of last resort.[223]

Trial court records, however, are a different story. It is unlikely that our findings are generalizable to the records held by trial courts. Unlike appellate courts, which are primarily engaged in deciding questions of law, trial courts must resolve competing factual claims. As a result, their files likely contain a higher frequency of sensitive information from a wider array of records,[224] including pre-trial discovery materials, expert reports, juror questionnaires, court transcripts, physical evidence, and audiovisual materials.[225] Nevertheless, the methods we used in coding and analyzing the North Carolina Supreme Court's records could be applied to trial court records as well, and we hope that other researchers will do so.

222. *See supra* note 161 and accompanying text. We did not review the "record on appeal," which is a separate filing containing, *inter alia*, copies of the case pleadings, jury instructions, transcripts, and other evidence entered in the lower courts.

223. Unfortunately, we could not find any data comparing state court appellate caseloads during the time period we studied. The National Center for State Courts (NCSC) did not implement the *State Court Guide to Statistical Reporting* until 2007, *see supra* note 180, and has not reclassified its historical data using the new reporting schema outlined in the guide. *See* Email to David Ardia from Shauna M. Strickland, Senior Court Research Analyst, NCSC, dated July 1, 2015 (on file with authors). Because of this, we cannot match our results with the data reported by other states during the time period we studied. We hope that other researchers will collect the data needed to do such comparisons.

224. We are aware of no comprehensive studies of trial court records. Carl Malamud, who used software to search for unredacted social security numbers in federal district court filings on PACER, noted that "often when our tool reported a Social Security number violation, when we looked around the document we also picked up many other Social Security numbers, birth dates, driver license numbers, Alien IDs, and bank account numbers." Malamud Letter, *supra* note 6, at 1.

225. Some of these materials may end up in an appellate court's files if they are submitted as part of the record on appeal, but rarely do they appear to the same extent in the parties' briefs and appendices. *See* Conley et al., *supra* note 13, at 776 (noting that trial court records "contain an abundance of personal information, some of which may drop away as cases move from trial courts to appellate courts").

A. IDENTIFYING WHERE PRIVACY RISKS ARE GREATEST

We knew going into this study that court records contain sensitive information. Indeed, scholars have long argued that court records raise substantial privacy concerns. What we lacked, however, was comprehensive data about the extent and context of this information. These data are essential to understanding the threats to privacy that court records present.

In Part V, we identified the most common types of sensitive information that appear in the North Carolina Supreme Court's case files. In this section we begin to evaluate the potential "harmfulness" of this information, based on frequency of appearance, certain document and case attributes, and existing legal authority requiring or recommending redaction. We note at the outset that "harmfulness" is a contested concept in privacy law,[226] and we do not take a position in this Article as to how "harm" should be defined in the context of public records. Instead, we present an assessment of relative risk based on the frequency of occurrence and context of a broad range of sensitive information types. Regardless of how one defines the harm that comes from the disclosure of certain types of sensitive information, the harmfulness of that information will likely be influenced by the frequency of disclosure and its context both within court records and the larger information ecosystem.

The following sections highlight our key findings.

1. Court Records Vary Substantially in the Sensitive Information They Contain

We found that court records vary substantially in both the types and frequency of sensitive information they contain. The records we studied did not exhibit every type of sensitive information in equal measure. Some information appeared much more often than other information. And some types of information that privacy advocates have highlighted did not appear at all in the records we studied. Moreover, nearly one in ten documents in our sample did not contain any of the 140 sensitive information types we coded for in this project.

226. As we noted in Part III, there is considerable disagreement among privacy scholars about the nature of the interests that privacy advances. In fact, some scholars argue that courts should abandon harm-based rationales entirely when evaluating privacy claims. *See, e.g.*, James Peterson, *Behind the Curtain of Privacy: How Obscenity Law Inhibits the Expression of Ideas About Sex and Gender*, 1998 WIS. L. REV. 625, 632 n.38 (1998) ("In practice, the distinction between *harm* and *offense* is not always easy to maintain, because extreme forms of offense can cause emotional and even palpable physical harm.").

Although it is not surprising that certain types of sensitive information appeared more often than others, we were surprised by some of the patterns we found. To facilitate the collection and analysis of our data, we grouped the various information types into thirteen categories.[227] When we compared these categories, we found that sensitive information in seven categories appeared much more frequently than the other categories of information. The most commonly occurring categories—location, identity, criminal proceedings, health, assets, financial information, and civil proceedings—each appeared in at least 20% of the documents we studied; the remaining six categories each appeared in fewer than 8% of the documents.[228]

Information types in the less frequently appearing categories warrant comment because of their unexpected absence in the documents. Information about sexual practices appeared infrequently, as did information in the intellectual pursuits category, which includes religious beliefs, political opinions, and voting and reading records.[229] Information about education was also mostly absent from the documents as were photos and videos. None of the documents contained any sensitive information that fell within the computer use category (e.g., user names, passwords, and search history).

We also found substantial variability in how often certain types of sensitive information appeared in individual documents. For most of the information categories, sensitive information appeared between three and six times per document.[230] There were outliers, however. Information in the criminal proceedings category appeared far more frequently in the documents than any other information category, showing up approximately nine to eighteen times as often as the other categories.[231] Interestingly, although information in the images and sexual activities categories did not appear in very many documents, when it did appear, it appeared more often in a document than all the other categories except the criminal proceedings category.[232]

227. For a description of the categories we utilized, *see supra* Section III.B.

228. *See supra* Table 4 and accompanying text.

229. Intellectual pursuits are a category of information that only recently began to receive protection under U.S. privacy laws, but is the subject of an increasing amount of privacy scholarship. *See supra* Section III.B.11.

230. *See supra* Table 4 and accompanying text.

231. The "images" category is the other outlier, with a median of seventeen appearances per document. *See supra* Table 4.

232. *See supra* Section V.A.2 and accompanying text.

Again, it is just as important to focus on what we did not find in the records. The types of information that most people would associate with financial fraud and identity theft appeared less often than we expected. We found surprisingly few social security numbers, bank account numbers, credit card numbers, and other financial account numbers, each of which appeared in no more than three documents in our sample of 504 documents.[233] We also found no partial social security numbers, debit card numbers, credit reports, or personal identification numbers (PINs) in the records.

Several types of sensitive health information also appeared less frequently than we expected. Information about abortion, paternity, and communicable diseases each appeared in only one document. Genetic information appeared in only three documents, which may be due in part to the time period we studied (1984–2000),[234] and information about drug or alcohol treatment appeared in only four documents. We found no references in the documents to an identified individual having HIV or AIDS.

2. Criminal Information Is Pervasive in Court Records

What we did find in great numbers in the court records was information related to criminal proceedings, particularly witness names, crime victim names, arrests, criminal charges, and the names of subjects under investigation. Indeed, information in the criminal proceedings category pervaded the court records we reviewed. It not only appeared in most of the documents, it also appeared more often in those documents than any other category of sensitive information.

More than half of the documents we analyzed contained information that fell in the criminal proceedings category, and the individual information types in this category far outpaced every other type of information we coded for in terms of frequency of occurrence. The names of witnesses in criminal cases appeared more often than any other information type in our data set, followed by the name of an individual

233. *See supra* Figure 3 and accompanying text. Although the North Carolina Rules of Appellate Procedure do not require that social security numbers be excluded from briefs filed in the North Carolina Supreme Court, the rules do state that SSNs "shall be deleted or redacted from any document before including the document in the record on appeal." N.C. R. APP. P. 9(a)(4). Our project did not involve the coding of the records on appeal.

234. *See* MICHAEL LYNCH ET AL., TRUTH MACHINE: THE CONTENTIOUS HISTORY OF DNA FINGERPRINTING 13 (2008) (noting that DNA evidence was first used in criminal investigations in the late 1980s but challenges in the courts and scientific press slowed its acceptance until the mid-1990s).

who was the subject of a criminal investigation. The name of a victim of criminal activity other than rape (rape victim name is a separate information type) was the third most frequently occurring information type, and information about arrests and convictions were the fourth and fifth most common information types, respectively.

Additionally, when information about criminal proceedings appeared in a document, it did so in large numbers. As we noted above, for most of the information categories sensitive information appeared between three and six times per document. By comparison, information in the criminal proceedings category appeared more than fifty times per document.[235] This substantially higher frequency of appearance is not due to a profusion of criminal cases. Documents from criminal cases made up only 36% of our sample.[236] Instead, what we see in the data is that criminal information appeared in documents filed in every type of case, including civil and juvenile cases.

Our data cannot tell us why criminal information is so pervasive in court records, but we can speculate based on qualitative factors. Information types in the criminal proceedings category—and in the civil proceedings category as well—relate to the functioning of the court system itself,[237] so it is perhaps not surprising to find these information types throughout the North Carolina Supreme Court's records. Indeed, when we examine the context of information from both of these categories, we see that information about the functioning of our criminal and civil courts is widely dispersed across all of the document types and case types we coded. As we noted, however, our data cannot definitely answer why criminal information is so common in court records. We hope that other researchers will take up this question.

3. Criminal Cases Have Disproportionately More Sensitive Information

Although documents from criminal cases constituted only slightly more than a third of the sample, they had an outsized impact on the types and frequency of sensitive information. More than three quarters (76.3%) of the sensitive information came from documents filed in criminal cases.

235. The criminal category appeared at a median frequency of appearance per document of 54.5. *See supra* Table 4 and accompanying text.

236. *See supra* Table 2.

237. Briefs filed in appellate courts often include arguments that turn on how the justice system functions, including acts and omissions by law enforcement, the credibility of witnesses, the relevancy of prior convictions and civil judgments, and the fairness of the jury.

And not only did criminal cases contain more sensitive information than civil and juvenile cases, they also contained a greater variety of sensitive information.

Criminal cases contained sensitive information from all of the categories that we identified in the documents,[238] whereas information from two categories, education and images, was entirely absent from civil cases.[239] Overall, criminal cases contained more individual types of sensitive information than civil and juvenile cases. Of the 140 sensitive information types that we coded for in the records, ninety-five distinct types appeared in the documents. Although some types appeared exclusively or nearly so in civil cases, documents filed in criminal cases contained a greater number of sensitive information types.[240]

Criminal cases also had substantially more sensitive information per document than either civil or juvenile cases. In fact, the median frequency of sensitive information in documents filed in criminal cases was approximately five times greater than that of documents filed in either civil or juvenile cases.[241] As a result, criminal cases had a higher density of sensitive information per page than either civil or juvenile cases. As page length increased, the number of pieces of sensitive information in criminal cases increased at a higher rate than it did in civil and juvenile cases. For criminal cases, the ratio between page length and frequency of sensitive information was four times greater than that of civil and juvenile cases.

One of the drivers of this disparity may be the disproportionate impact of death penalty cases.[242] Although documents from death penalty cases constituted only 6.5% of our sample,[243] they contained more than a quarter

238. We did not find any information in the documents that fell within the "computer use" category. *See supra* Table 4.

239. Six of the thirteen information categories were absent from documents filed in juvenile cases.

240. *See supra* Table 5 and accompanying text.

241. Sensitive information in criminal cases also appeared much more frequently in the brief body than in the appendix. *See supra* Figure 2.

242. During the period from 1989–1998, there was a sharp increase in the number of documents from cases in which a defendant was sentenced to death (from 0.8 documents per year to 2.7 per year). This corresponded to an increase in death penalty verdicts in North Carolina, which rose from nine in 1989 to thirty-four in 1995. CTR. FOR DEATH PENALTY LITIG., ON TRIAL FOR THEIR LIVES: THE HIDDEN COSTS OF WRONGFUL PROSECUTIONS IN NORTH CAROLINA 16 (2015), http://www.cdpl.org/wp-content/uploads/2015/06/INTERACTIVE-CDPL-REPORT.pdf. In North Carolina, death penalty convictions are subject to automatic review by the North Carolina Supreme Court. N.C. GEN. STAT. ANN. § 15A-2000(d)(1) (West 2015).

243. The most commonly occurring appellate subject area was "tort, contract, and real property," which constituted 32% of the documents in the sample, followed by

(27.7%) of the total amount of sensitive information. Documents from death penalty cases also were, on average, the longest documents in the sample.[244] Interestingly, most of the appearances of sensitive information types in the education category occurred in documents filed in death penalty cases, whereas information types in the employment and financial categories were largely absent from documents filed in these cases.[245]

4. Minors Deserve Additional Attention

The vast majority of the sensitive information was associated with adults. Nevertheless, because of heightened concerns about the privacy of children, information associated with minors deserves special attention.[246]

As the data showed, sensitive information about minors was not limited to juvenile cases. Overall, 7% of the sensitive information was associated with an identified minor. Criminal cases evidenced slightly more information associated with minors than civil cases, but the difference was not substantial.[247] Not surprisingly, juvenile cases contained significantly more information about minors: 40% of the sensitive information in juvenile cases was associated with a minor.[248]

"felony (non-death penalty)," which arose in 26% of the documents. *See supra* notes 180–181 and accompanying text.

244. Documents filed in death penalty cases had a mean length of 113.63 pages, more than twice the average length for all documents in the sample (sample mean was 47.92 pages).

245. The higher frequency of appearance of information relating to education in death penalty cases may be due to its relevance as mitigation evidence. *See* N.C. GEN. STAT. § 15A-1340.16(e). Documents in death penalty cases also had the highest incidences of the following sensitive information types: "Criminal Sentence," "Incarceration," "Juror Names," and "Student Discipline."

246. Some commentators have argued that all juvenile proceedings should be closed by default in order to protect the interests of minors. *See, e.g.*, William Wesley Patton & Kelly Crecco, *An Update to Striking a Balance: Freedom of the Press Versus Children's Privacy Interests in Juvenile Dependency Proceedings*, 12 FIRST AMEND. L. REV. 575, 589 (2014) ("Children are at much more risk of juragenic psychological harm in a presumptively open dependency court system than in a discretionarily open court system for a number of reasons."); *see also* Kristin Henning, *What's Wrong with Victims' Rights in Juvenile Court?: Retributive Versus Rehabilitative Systems of Justice*, 97 CALIF. L. REV. 1107, 1158–60 (2009) (suggesting there can be tensions between victim rights and the confidentiality of juvenile prosecutions).

247. Sensitive information in criminal cases was associated with minors 7.2% of the time. In civil cases it was 6.1%.

248. *See supra* Section V.B.2.b. Although we did not record whether the minor in question was a "covered juvenile" under North Carolina law and thus entitled to additional privacy protections, a recent study published by the Juvenile Law Center found that the vast majority of states—including North Carolina—are failing to protect highly sensitive information contained in juvenile court records. JUVENILE LAW CTR., FAILED

Although the overall amount of information associated with minors was low, there were some notable differences between the information categories with regard to minors. Relative to their baseline percentage across all categories, minors were less likely to be associated with information relating to criminal proceedings and more likely to be associated with information in the education, health, identity, and sexual activities categories.[249] In addition, a number of specific information types were disproportionately associated with minors. Information about "Communicable Disease" and, unsurprisingly, "Name of Minor Child," for example, were exclusively identified with minors,[250] and several information types associated with both adults and minors appeared more often with minors than the overall percentages would have suggested, including: "Adoption," "Custody or Guardianship," "Date of Birth," "Juvenile Court History," "Pregnancy," and "Sex Life."

On the other hand, there were several information types we expected to find associated with minors more often than we observed. For instance, no photographs or videos were associated with minors. Information about "Student Discipline" appeared in only one document, and "Student Grades and Performance Evaluations" appeared in only six documents.

Given the small number of documents from juvenile cases in the sample (n = 7), the inferences regarding the extent of sensitive information associated with minors is limited, especially in juvenile cases. A larger sample of documents from juvenile cases is necessary to better understand the privacy risks associated with minors. This is also an area we hope future researchers will explore.

5. It Is Unwise to Focus Exclusively on Appendices

When we began this study we assumed, based largely on anecdotal reports from archivists and privacy scholars, that appendices included in court records would contain more sensitive information than legal briefs and that highly sensitive information that had been kept out of legal briefs would nevertheless appear in the appendices. Our study showed, however, that appendices are for the most part not quantitatively different from

POLICIES, FORFEITED FUTURES: A NATIONWIDE SCORECARD ON JUVENILE RECORDS (2015), http://jlc.org/blog/new-study-reveals-majority-us-states-fail-protect-juvenile-records.

249. *See supra* Section V.B.2.b.

250. Information concerning "Communicable Disease" appeared only once in the sample. As noted in Part III, we coded for "Name of Minor Child" as a specific information type; this was the most common information type associated with minors. *See supra* Section V.B.2.b, especially Figure 7.

legal briefs in terms of the frequency and types of sensitive information they contain.

In terms of the amount of sensitive information in a document, the data actually showed that legal briefs contained a higher frequency of sensitive information than appendices. This disparity was particularly evident in criminal and juvenile cases, where the brief bodies contained approximately three times as much sensitive information as the appendices; for civil cases, sensitive information appeared with equal frequency in the appendices and briefs.[251] These findings were reinforced by the multiple regression model, which showed that a document's inclusion of an appendix did not affect the total amount of sensitive information in the document.[252]

With regard to the specific types of sensitive information in briefs and appendices, the results were more mixed. Only seven of the ninety-five information types that we identified in the records appeared more often in the appendices.[253] On the other hand, twenty-seven information types appeared exclusively in the briefs, and many more appeared more than 50% of the time in the briefs. Nevertheless, the information types that did appear more often in the appendices were the types of information many would regard as particularly sensitive from the standpoint of identity theft. Uniquely identifying physical characteristics, drivers' license numbers, social security numbers, and state identification numbers all appeared more often in the appendices, with the latter three information types appearing exclusively in an appendix.

Accordingly, although there is good reason to pay careful attention to appendices when reviewing court records for sensitive information, it is unwise to focus exclusively on appendices. More types of sensitive information appear in legal briefs, and at a higher overall frequency than in appendices.

6. Trends in Sensitive Information over Time

Although the amount of sensitive information in the case files of the North Carolina Supreme Court varied significantly during the time period of this study (1984–2000), there were no overarching trends in the frequency of sensitive information during this seventeen-year period.[254]

251. *See supra* Figure 2 and accompanying text.
252. *See supra* Section V.B.4.
253. *See supra* Figure 8 and accompanying text.
254. *See supra* Figure 10 and accompanying text.

Instead, the most commonly occurring information types appeared with a frequency that varied within a relatively consistent range.

This is not to say that the amount of sensitive information was constant during this time period. To the contrary, there was a great deal of year-to-year variability. In some years there were sharp increases in the amount of sensitive information while in other years there were steep declines. Moreover, the various categories of information did not rise and fall together.

Nevertheless, these variations, which appear to be cyclical, do not show a declining or rising trend during the time period under study. Whether this would continue to be the case if we extended our collection of court records earlier or later in time is beyond the scope of this study.

A number of important events occurred in the late 1990s and early 2000s that might have a significant impact on the frequency and extent of sensitive information in court records in the years following our study. In 1999, the North Carolina court system began allowing electronic filing and in 2009 implement e-filing rules that placed the onus on the parties to redact a number of types of sensitive information from case filings, including social security numbers and certain financial information.[255] In addition, in the late 1990s there was a significant rise in the use of computers and electronic communication systems that might have led to the generation of different types of sensitive information in court records. Given the time it takes for a case to work its way up to a state's highest court, we can expect that these changes would likely take a few years to be reflected in the North Carolina Supreme Court's files.

B. CHALLENGES IN IMPLEMENTING PRIVACY PROTECTIVE PRACTICES

In addition to aiding our understanding of privacy in the context of court records, the results of this research and the experience of the coders will have practical implications for court personnel and archivists as they develop rules and practices for electronic filing of court records or the digitization of older records. Although all of our findings should have some bearing on these efforts, we highlight four main points.

First, some types of sensitive information are easier to identify in court records than others. If courts or archivists decide to limit access to certain

255. *See generally* Deborah Leonard Parker, *Electronic Filing in North Carolina*, 2 J. APP. PRAC. & PROCESS 351 (2000) (describing the impact of the introduction of electronic filing rules in North Carolina in 1999). *See also supra* note 10 and accompanying text.

types of sensitive information, they will inevitably have to make choices about which information types should be restricted based, at least in part, on whether the burden of addressing privacy is commensurate with the risk of harm and whether investments in privacy protection practices will be effective. Such proportionality considerations get to the heart of the debate about the nature of privacy harms and the risks, both foreseeable and perhaps unforeseeable, that are presented by current and future practices of information aggregation and use. Without solving the normative and broader practical problems, though, this study nonetheless reveals that some of the sensitive information types in court records appear in standard formats—such as social security numbers, dates of birth, and financial account numbers—and therefore are more easily identifiable through automated searching techniques.[256]

Second, many sensitive information types require human readers to review records more than once in order to identify the information. Our coders reported that some sensitive information was identifiable only by reading the record and developing a sense of the narrative. For example, the first mention of an individual might not reveal that she was indeed a cooperating defendant or that a particular named person was a minor. This type of information would likely require a human reader to review the record and make note of names associated with sensitive information. Coders also reported that reading a document once would not necessarily be sufficient to capture all occurrences of names that could be associated with sensitive information. Even after this investment of human effort, the question remains about how best to balance privacy and judicial transparency if the sensitive information arises through the course of the narrative. The name itself might be amenable to redaction, but the story of cooperation might be too interwoven in a brief or appendix to make redaction feasible.

Third, redaction may be a poor strategy for dealing with some sensitive information types. Our coders reported that some court briefs and

256. *See* Rebecca Green, *Petitions, Privacy, and Political Obscurity*, 85 TEMP. L. REV 367, 406 n.272 (2013) (noting that courts and judicial administrators have explored redaction using computer software, but also noting the prevalence of redaction errors in the federal court records PACER database); Ronald Leighton, Joe Cecil, Michael Ishakian & Edward Felten, *Panel Three: Implementation—What Methods, If Any, Can be Employed to Promote the Existing Rules' Attempts to Protect Private Identifier Information from Internet Access?* 79 FORDHAM L. REV. 45, 49 (2010) (Felten discusses the amenability of social security numbers to software redaction because of their fixed pattern and suggests that advanced machine learning methods could be developed to help locate and redact even "difficult types of information, such as names of minor children.").

accompanying appendices revealed information that was difficult to code as a discrete occurrence. The story of child abuse, for example, might pervade a particular case record. Options for addressing privacy would need to account for this challenge. The preliminary coding also revealed that the occurrence of language conveying a person's gender is a poor fit for redaction because gender is so integral to the English language. As explained in Part III, the coding for gender threatened to overwhelm the coding process, and the same would be true for any redaction effort.

Fourth, the data showed that there might be some value in prioritizing the review of certain documents when searching for sensitive information in appellate court records. We found that the various brief types were not all equal in the amount of sensitive information they contained. For example, briefs filed by the state had the highest frequency of sensitive information, whereas amicus briefs had the fewest appearances of sensitive information.[257] In addition, documents filed in death penalty cases had a disproportionately higher rate of sensitive information than other types of cases. For court personnel and archivists seeking to make the best use of limited resources, it may make sense to focus on some types of documents and cases over others. We would caution, however, against focusing exclusively on appendices.[258]

VII. CONCLUSION

Court records present a special challenge for privacy advocates. Unlike in many other areas of privacy law, the information in court records is presumptively open to the public. This openness serves many salutary functions, such as ensuring that our system of justice functions fairly and is accountable to the public. The public's right of access to court records and the information they contain, however, is not absolute.

Courts can—and frequently do—restrict public access when an overriding interest supports closure or sealing of specific information. Although the precise standard that a court must apply will vary depending on the circumstances, in general courts must conclude that the interest in prohibiting disclosure outweighs the strong presumption of public access. In the context of libraries and other archives, which may not be bound by

257. Briefs filed by the state had a median of 105.5 appearances of sensitive information per document; for amicus briefs, the median was 10.5. *See supra* Table 3. Multiple regression modeling also showed that the brief types (other than amicus briefs) were statistically significant predictors of the amount of sensitive information in a document.

258. *See supra* Section VI.A.5.

law to provide public access to court records, the question is not about what the law requires but about what policy best ensures the protection of privacy interests while simultaneously informing the public about the functioning of the court system.

Although we found a substantial amount of sensitive information in the court records we studied, we have not sought to tell courts or archivists what information, if any, should be redacted or what documents should be withheld from online access or otherwise managed for privacy protection. These largely normative questions must be answered based on a careful balancing of the competing public access and privacy interests. The data presented in this study will be helpful in this balancing, but the findings should not be read to dictate one approach or another.

Privacy interests cannot be evaluated based solely on the presence or absence of specific types of sensitive information in a single court document. Other factors, including the context of the information and the extent of information about an individual that is available from other sources, must also be taken into account. On that latter point, this study is but one piece in a complicated mosaic.

What this study has shown is that court records vary significantly in the types of sensitive information they contain. Records in civil cases are not identical to records in criminal cases or juvenile cases. Consequently, when scholars and policymakers discuss privacy and court records, they must be cautious of generalizing. Depending on the privacy concerns considered paramount, we are likely to see a very different risk profile between different types of court records, cases, and levels of the court system.

Much work remains in order to understand the privacy risks that might arise from online access to court records. We hope that future researchers will answer some of the questions that the data have raised, including the prevalence of criminal information in court records, the differences between appellate court records and trial court records, the extent of sensitive information in juvenile court files, and the impact of e-filing procedures on the types and frequency of sensitive information in court records.

APPENDIX

Table A1: Sensitive information types in coding list, with category and number of documents that contained each information type, percentage of the total sample of documents (*n* = 504), overall frequency of appearance, and percentage of total frequency of all sensitive information identified in the sample.

		Documents		Frequency	
Sensitive Information Type	**Category**	***n***	**%**	***n***	**%**
Abortion	Health	1	0.2%	1	0.0%
Adoption	Civil Proceedings	4	0.8%	114	0.2%
Age	Identity	84	16.7%	350	0.7%
Arrest or Charge	Criminal Proceedings	160	31.7%	2,512	4.7%
Asset-Other	Assets	3	0.6%	6	0.0%
Bank Account Number	Financial	3	0.6%	9	0.0%
Bankruptcy	Financial	4	0.8%	10	0.0%
Cable Television Subscription Record	Intellectual Pursuits	1	0.2%	4	0.0%
Cable Television Viewing History	Intellectual Pursuits	0	-	0	-
Cause of Death	Health	70	13.9%	174	0.3%
Child Abuse	Criminal Proceedings	27	5.4%	378	0.7%
Child Support	Civil Proceedings	11	2.2%	40	0.1%
Civil Commitment	Civil Proceedings	3	0.6%	27	0.1%
Civil Proceedings-Other	Civil Proceedings	2	0.4%	5	0.0%
Communicable Disease	Health	1	0.2%	1	0.0%
Compensation	Financial	125	24.8%	909	1.7%
Computer Use - Other	Computer Use	0	-	0	-
Content of Recorded Conversations	Intellectual Pursuits	3	0.6%	119	0.2%
Conviction	Criminal Proceedings	150	29.8%	1,744	3.3%
Cooperating Defendant Name	Criminal Proceedings	32	6.3%	772	1.5%
Credit Card Number	Financial	1	0.2%	2	0.0%
Credit Report	Financial	0	-	0	-
Criminal Proceedings-Other	Criminal Proceedings	17	3.4%	257	0.5%
Custody or Guardianship	Civil Proceedings	17	3.4%	257	0.5%
Date of Birth	Identity	30	6.0%	46	0.1%
Date of Death	Identity	85	16.9%	325	0.6%
Date of Hospital Stay	Health	37	7.3%	115	0.2%
Debit Card Number	Financial	0	-	0	-
Debt	Financial	7	1.4%	47	0.1%
Dependency or Neglect	Civil Proceedings	2	0.4%	15	0.0%
Disability Status	Health	24	4.8%	136	0.3%

		Documents		Frequency	
Sensitive Information Type	**Category**	***n***	**%**	***n***	**%**
Discipline	Employment	20	4.0%	237	0.5%
Divorce	Civil Proceedings	35	6.9%	200	0.4%
Domestic Violence Victim Name	Criminal Proceedings	7	1.4%	525	1.0%
Driver's License Number	Identity	2	0.4%	6	0.0%
Drug or Alcohol Dependency	Health	34	6.7%	325	0.6%
Drug or Alcohol Treatment	Health	4	0.8%	6	0.0%
Drug or Alcohol Use	Health	8	1.6%	135	0.3%
Education-Other	Education	1	0.2%	4	0.0%
Eligibility for School Lunch Program	Education	0	-	0	-
Email Address	Identity	0	-	0	-
Employment-Other	Employment	0	-	0	-
Fax Number	Identity	0	-	0	-
Financial Aid Award	Education	0	-	0	-
Financial Asset	Assets	74	14.7%	492	0.9%
Financial-Other	Financial	5	1.0%	11	0.0%
Fingerprint	Identity	0	-	0	-
Foreclosure Judgment	Financial	6	1.2%	81	0.2%
Full-Face Photograph	Images	1	0.2%	2	0.0%
Gait	Identity	0	-	0	-
Gender Identity Change	Identity	0	-	0	-
Genetic Information	Health	3	0.6%	8	0.0%
Geolocation Information	Location	52	10.3%	301	0.6%
Gun Permit	Assets	0	-	0	-
Gun Permit Application	Assets	0	-	0	-
Gun Possession or Ownership	Assets	65	12.9%	781	1.5%
Health Plan Beneficiary Number	Health	0	-	0	-
Health-Other	Health	65	12.9%	781	1.5%
HIV / AIDS Status	Health	0	-	0	-
Home Address	Location	55	10.9%	216	0.4%
Identity-Other	Identity	13	2.6%	32	0.1%
IM ID	Computer Use	0	-	0	-
Images-Other	Images	0	-	0	-
Incarceration	Criminal Proceedings	49	9.7%	295	0.6%
Informant Name	Criminal Proceedings	4	0.8%	290	0.6%
Insurance Policy Number	Financial	5	1.0%	14	0.0%

		Documents		Frequency	
Sensitive Information Type	**Category**	***n***	**%**	***n***	**%**
Intellectual Pursuits-Other	Intellectual Pursuits	1	0.2%	1	0.0%
Internet Protocol (IP) Address	Computer Use	0	-	0	-
Internet Search History	Computer Use	0	-	0	-
ISP Records	Computer Use	0	-	0	-
Iris Print	Identity	0	-	0	-
Juror Name	Criminal/Civil Proceedings	35	6.9%	1,563	3.0%
Juvenile Court History	Criminal Proceedings	4	0.8%	19	0.0%
Loan Account Number	Financial	0	-	0	-
Location-Other	Location	101	20.0%	1,122	2.1%
Medical Billing Number	Health	0	-	0	-
Medical Condition	Health	101	20.0%	1,122	2.1%
Medical Device ID/Serial Number	Health	0	-	0	-
Medical Record Number	Health	0	-	0	-
Military Discharge	Employment	8	1.6%	11	0.0%
Mother's Maiden Name	Identity	5	1.0%	9	0.0%
Mug Shot	Criminal Proceedings	0	-	0	-
Name of Minor Child	Identity	85	16.9%	1,581	3.0%
Name of Subject of Investigation	Criminal Proceedings	173	34.3%	8,284	15.6%
Other Crime Victim Name	Criminal Proceedings	136	27.0%	4,988	9.4%
Other Financial Account Number	Financial	2	0.4%	4	0.0%
Other Health Diagnosis or Treatment	Health	99	19.6%	1,397	2.6%
Parole Status	Criminal Proceedings	5	1.0%	24	0.1%
Passport Number	Identity	0	-	0	-
Password	Computer Use	0	-	0	-
Paternity Test	Health	1	0.2%	1	0.0%
Performance Evaluation	Employment	16	3.2%	59	0.1%
Personal Identification Code/Password	Financial	0	-	0	-
Photos/Videos - Fully Undressed	Images	0	-	0	-
Photos/Videos - Partially Undressed	Images	0	-	0	-
Photos/Videos -Violence, Abuse, Death	Images	1	0.2%	15	0.0%
Place of Birth	Identity	6	1.2%	7	0.0%
Place of Death	Health	28	5.6%	41	0.1%
Political Opinion	Intellectual Pursuits	2	0.4%	9	0.0%
Pregnancy	Health	14	2.8%	56	0.1%

		Documents		Frequency	
Sensitive Information Type	**Category**	***n***	**%**	***n***	**%**
Prescriptions	Health	7	1.4%	17	0.0%
Presentence Investigation Report	Criminal Proceedings	1	0.2%	1	0.0%
Prior Civil Judgment	Civil Proceedings	60	11.9%	528	1.0%
Professional Cert. or License Number	Identity	29	5.8%	76	0.1%
Racial or Ethnic Origin	Identity	19	3.8%	175	0.3%
Rape Victim Name	Criminal Proceedings	43	8.5%	1,121	2.1%
Real Estate Ownership/Rental	Assets	82	16.3%	1,103	2.1%
Records of Library Use	Intellectual Pursuits	0	-	0	-
Records of Reading Material	Intellectual Pursuits	3	0.6%	5	0.0%
Religious or Philosophical Belief	Intellectual Pursuits	16	3.2%	172	0.3%
RFID	Computer Use	0	-	0	-
School Address	Location	0	-	0	-
Sentence	Criminal Proceedings	3	0.6%	5	0.0%
Sex Life	Sexual Activities	16	3.2%	172	0.3%
Sex Video	Sexual Activities	0	-	0	-
Sexual Abuse Allegation	Criminal Proceedings	3	0.6%	12	0.0%
Sexual Activities-Other	Sexual Activities	129	25.6%	776	1.5%
Signature	Identity	0	-	0	-
SSN - Full	Identity	1	0.2%	1	0.0%
SSN - Partial	Identity	0	-	0	-
State ID Number	Identity	1	0.2%	1	0.0%
Student Discipline	Education	1	0.2%	1	0.0%
Student Grades or Performance Evals.	Education	6	1.2%	16	0.0%
Student ID	Education	0	-	0	-
Tax Lien	Financial	5	1.0%	5	0.0%
Tax Return	Financial	1	0.2%	5	0.0%
Telephone Number	Identity	231	45.8%	570	1.1%
Trade Union Membership	Intellectual Pursuits	0	-	0	-
Unique Physical Characteristic	Identity	5	1.0%	27	0.1%
User Name	Computer Use	0	-	0	-
Vehicle Identification Number	Assets	0	-	0	-
Vehicle License Plate Number	Assets	2	0.4%	3	0.0%
Victim of Identity Theft	Financial	0	-	0	-
Video Rental Records	Intellectual Pursuits	1	0.2%	1	0.0%
Voice Print	Identity	0	-	0	-

		Documents		Frequency	
Sensitive Information Type	**Category**	***n***	**%**	***n***	**%**
VOIP ID	Computer Use	0	-	0	-
Voting Record	Intellectual Pursuits	1	0.2%	2	0.0%
Witness Name	Criminal/Civil Proceedings	221	43.8%	14,437	27.2%
Work Address	Location	253	50.2%	752	1.4%
Zip Code	Location	202	40.1%	853	1.6%

Table A2: Regression results for the analysis of sensitive information per document (log transformed).

Variable	**Coefficient**	**Standard Error**
Criminal Case Type	1.552622*	0.1512302
Juvenile Case Type	0.3569533	0.4766879
Amicus Curiae Brief	1.185187	0.9326415
Appellant's Brief	2.32089*	0.8866871
Appellee's Brief	1.920483*	0.8927833
Petition for Discretionary Review	1.951337*	0.8965791
Brief for the State	2.237307*	0.8981081
Appendix Present	0.0542404	0.1371782
Document Length (in pages)	0.0109019*	0.0011283
Constant	0.2973103	0.8927754
R-squared	0.4626	
No. of observations	467	

* indicates statistical significance ($P < 0.05$)