

American Courtroom Data: A Legal Guide for Researchers

David Griffin ¹ and Alex Moe^{†2}

¹Aston University, david@linguistlawyer.com

²Loyola University Chicago School of Law

Abstract

Subject to very few exceptions, courtroom data generated in the United States (i.e. any material that is either part of a docket or derived from an official courtroom event) is part of the public domain. Once a researcher has acquired such data, unless it is subject to separate preexisting copyright or other similar protections, there are no legal restrictions on its use. This is true at both the state and federal levels, and regardless of the citizenship of the researcher in question. Courtroom data can always be acquired by visiting the clerk of court's office in which it has been filed and can increasingly be obtained online as well. When working with courtroom data from the United States, researchers are advised to consider the presence of personally identifying information in that data and whether it should be obscured or anonymized prior to the publication of any results.

Keywords: *US courts, US law, courtroom data, legal documents, ethics.*

Resumo

Salvo raras exceções, os dados provenientes das salas de audiências dos Estados Unidos (ou seja, qualquer material que faça parte de um registro ou seja derivado de um evento oficial da sala de audiências) fazem parte do domínio público. Assim que um investigador tenha acesso a esses dados, a menos que estejam sujeitos a direitos de autor pré-existentes ou outras proteções semelhantes, não há restrições legais para a sua utilização. Isto aplica-se tanto a nível estatal como federal e independentemente da nacionalidade do investigador em questão. Os dados da sala de audiências podem sempre ser obtidos através de uma ida à secretaria do tribunal em que foram arquivados e, cada vez mais, podem também ser obtidos online. Ao trabalhar com dados de tribunais dos Estados Unidos, os investigadores devem ter em consideração a possível presença de informações pessoais e identificativas de indivíduos nesses dados e, ainda, se estas devem ser ocultadas ou anonimizadas antes da publicação de quaisquer resultados.

Palavras-chave: *tribunais norte-americanos, direito norte-americano, dados do tribunal, documentos legais, ética.*

[†]Alex Moe works for the Illinois Attorney General, and teaches legal writing and oral advocacy at Loyola University Chicago School of Law, but the views expressed herein are his alone.

1. Introduction

Courtroom data generated in the United States (referred to in this article as “American courtroom data”) is a frequent subject of analysis for forensic linguists, particularly for those at the undergraduate and postgraduate levels. There are a number of reasons for this popularity, not the least of which is the relative ease of its acquisition. In comparison, acquiring similar data from other common law legal systems like that of England and Wales can be prohibitively difficult. Despite this widespread use of American courtroom data in forensic linguistics, to the authors’ knowledge, there does not yet exist a formal guide written specifically for researchers without legal backgrounds that discusses in detail the regimes which regulate the acquisition and use of this material. The purpose of this article is therefore to provide such a guide. To begin with, this article will outline the key legal terms and concepts necessary to understand the various factors at play in this discussion. Afterwards, the article will explain the legal status of American courtroom data. It will then go on to discuss some of the ethical factors that should be considered when working with such data. To help readers distinguish between the legal and academic nature of the sources referenced in this article, statutes, case law, and other operative legal sources will be cited via footnotes while other sources (including legal sources cited for non-operative purposes) will be referenced via in-line citation. Throughout the article, readers should keep in mind that the legal discussion is strictly limited to the context of the US legal system. While the authors are not aware of any legal issues arising from the use of American courtroom data in other jurisdictions, such determinations should ultimately be made with the relevant qualified legal professionals.

2. Key Legal Terms and Concepts

In this article, “courtroom data” means any material that is either part of a docket (i.e. the official court record of the documents related to a given legal case) or directly derived from an official courtroom event (e.g. a stenographer’s transcript of a hearing or an audiovisual recording of a trial’s closing argument). Inasmuch as it covers data from across the whole of the judicial process regardless of modality, this definition of “courtroom data” may be more inclusive than what readers without legal training would expect. As will be discussed below, however, it is specifically the connection that this data has to a courtroom via either its being filed or its role in a given legal proceeding that determines its availability. To emphasize the essential nature of that connection, “courtroom data” is believed to be the most apposite term.

The contents of a case’s docket are generally referred to by lawyers as “filings”, regardless of whether a given document was originally written for the purpose of filing (e.g. a legal brief) or not (e.g. a piece of evidence). In line with that practice, the term “filing” will be similarly employed throughout this article. Legal materials that are generated as part of a lawsuit but not contained in a case’s docket are excluded from consideration here as such documents are generally covered by attorney-client privilege or otherwise not subject to public disclosure (Reagan, 2010, pp. 11-12).

Non-public filings, which are discussed at length in the following sections, are frequently referred to as having been “sealed”. While federal courts almost exclusively use

this term for such filings (Reagan, 2010, pp. 2-3),¹ some states have created multiple gradations of non-public status. The state of Illinois, for example, classifies non-public courtroom filings as impounded, confidential, sealed, or expunged (Administrative Office of the Illinois Courts, 2022a, p. 63),² with each layer denoting a particular level of availability. Impounded documents are available to the parties in a case but not the public, for example, while confidential documents are only available to the judge and the party who filed them. Regardless of the specific terminology employed, the only relevant consideration for this article is whether a given document is in fact publicly available, and hence “sealed” will be used to simplify the discussion.

Sealing is presumptively of indefinite duration, though some statutes permit for the unsealing of particular documents after certain conditions (e.g. the passing of a specified period of time) are met. While any person can technically request the unsealing of a given document, such a request is unlikely to be successful without substantial legal grounding. Due to the generally low probability of success of such requests and the variance in the procedures to do so across jurisdictions, the process of unsealing documents in this way will not be considered further here.

The legal system of the United States can be broadly divided into state and federal courts, with each level of court having distinct legal and jurisdictional competencies. Fortunately for present purposes, the legal and ethical considerations for working with documents from either a state or federal court are essentially identical. Throughout this article, references are provided to relevant federal statutes and case law, but for reasons of space, the presentation of the equivalent sources for all fifty US states is not possible. Because of the authors’ legal backgrounds, statutes and case law from the state of Illinois will be presented as exemplars. However, policies in the other 49 states vary only slightly. All legal references are up to date as of March 2025 and, as is discussed below, this is a very settled and uncontroversial area of US law; nevertheless, the authors encourage anyone particularly concerned to check the applicable state or federal materials for themselves.

Readers who are not particularly familiar with the legal system of the United States should be aware that despite its many similarities to other common legal systems, there are some marked differences in the terminology it employs. In a US court, for example, a “plaintiff” is the party who originally filed a civil suit, but this term has fallen out of use in many other legal systems. In the legal system of England and Wales, this party would instead be referred to as a “claimant”, while in Scotland, the term “pursuer” is used (Grainger & Fealy, 1999, p. 76; Incorporated Council of Law Reporting for England and Wales, 2022). In case of any confusion, readers are advised to consult the most recent edition of *Black’s Law Dictionary* (Garner, 2024a) or a similar reference work.

3. Legal Considerations When Using American Courtroom Data

The public right to inspect court documents in the United States’ legal system predates the country itself and is so well-established as to be “beyond dispute” (Ardia & Klinefelter, 2015, pp. 1817–1818).³ This right originally derives from the English common-law

¹Fed. R. Civ. P. 5.2.

²Ill. Sup. Ct. R. 8(b).

³*Leucadia, Inc. v. Applied Extrusion Technologies, Inc.* (1993). 998 F.2d 157, 161.

tradition (“The Common Law Right to Inspect and Copy Judicial Records: In Camera or On Camera”, 1982, p. 666), and can now be found in various Constitutional guarantees (most notably those of the First Amendment) (Ardia & Klinefelter, 2015, pp. 1819–1820), as well as a variety of federal and state laws governing record access and the administrative rules of individual courts (Lombard, 2006, pp.1069–1075; Ardia & Klinefelter, 2015, pp. 1819–1820).⁴ Tied up as it is with basic public rights, public access to court documents has proven a fertile ground for study and litigation for the past millennium (*The Common Law Right to Inspect and Copy Judicial Records: In Camera or On Camera*, 1982, pp. 666–671). For the researcher, however, the present-day takeaway is simple: American courtroom filings are presumptively public documents, available to any person, at any time, for any reason. While this rule is not absolute, it is nearly so. The following section explores the legal rules that govern access to the use of American courtroom data.

3.1. Confidentiality Regimes

Federal and state courts have broadly similar regimes governing legal disputes over public access to courtroom documents. Though a full examination of these regimes is not possible here, a brief review may be useful to the reader.

At the federal level, the starting point for modern jurisprudence regarding document access is *Nixon v. Warner Communications*, which concerned broadcasters’ attempts to access copies of the Watergate tapes.⁵ The US Supreme Court reaffirmed a presumption in favor of access to court records, and though the case was eventually resolved on other grounds, that presumption remains (*The Common Law Right to Inspect and Copy Judicial Records: In Camera or On Camera* 1982, p. 678).⁶ In a case where public access to a given courtroom filing is disputed, a court may apply either a common-law or First Amendment analysis, depending upon its jurisdiction (Reagan, 2010, pp. 3-4). Under the common law analysis, public access is presumed, and restrictions are judged by balancing the need for confidentiality in any given case with the public’s interest in open records (Reagan, 2010, pp. 3-4; Regalia, 2015, pp. 95-96). Under a First Amendment analysis, public access to courtroom documents is permitted where the type of proceeding that the document relates to has historically been open, and where keeping that proceeding open makes logical sense (Lombard, 2006, pp. 1071–1072).⁷

At the state level, the federal First Amendment analysis serves as a floor, meaning that no state can have an access regime more restrictive than that federal analysis requires, but states are allowed to build upon that floor to permit additional access in different ways. The leading case in Illinois, for example, is *Skolnick v. Alzheimer & Gray*, which rejected a request to file a fraud case under seal.⁸ Both the common law right of access and First Amendment analysis are applied in Illinois proceedings.⁹ Beyond that, Illinois has a statute guaranteeing public access to records that serves as an additional

⁴*Nixon v. Warner Comm’s.* (1978). 435 U.S. 589, 598.

⁵*Id.* at 598.

⁶*Id.* at 602.

⁷*Richmond Newspapers v. Va.* (1980). 448 U.S. 55.

⁸*Skolnick v. Alzheimer & Gray.* (2000). 191 Ill. 2d 214.

⁹*Id.* at 230–32.

protection for the right of public access to courtroom documents.¹⁰ The mechanics of the state analysis here do not differ notably from the federal analysis.

3.2. Sources of Confidentiality

Though it is nearly absolute, the presumption of public access to American courtroom data is not completely so. It is worth examining, therefore, the circumstances under which public access to a document can be restricted. There are three general reasons why access to a document may be restricted, which this section will explore in sequence: 1) the entire case to which a document belongs may be sealed; 2) certain types of document are required to be sealed; and 3) individual documents may be sealed on a case-by-case basis.

3.2.1. Sealed Cases

Under certain circumstances, an entire case may be sealed. Courtroom proceedings in a sealed case are not open to the public, related court filings are not available, and the name and even the very existence of such a case may be kept secret.¹¹ Such matters, however, are extraordinarily rare. At the federal level, sealed cases account for a very small fraction of all filings (Reagan & Cort, 2009, p. 4). Though it is more difficult to determine the frequency of sealed cases at the state level, based on the authors' analysis of the data in Administrative Office of the Illinois Courts (2022a), the overall percentage of filings that are restricted in this way is likely in the low single digits. The higher frequency of sealed cases in state court is based on the types of cases that are filed in each system, rather than a result of an inherent difference between how federal and state courts treat sealed cases.

The types of cases that are sealed in this way are defined by statute and the reasons for doing so tend to be relatively straightforward. By way of example, most case types related to juveniles—including child abuse proceedings, adoption, and juvenile delinquency cases—are subject to partial or total seal in perpetuity (Ardia, 2017, pp. 1430–1434; Administrative Office of the Illinois Courts, 2022b, pp. 64–66). The most common type of sealed case at the federal level is a *qui tam* action, where a private whistleblower essentially seeks to recover money that the government improperly paid as a result of fraud (Reagan & Cort, 2009, p. 5).¹² In a *qui tam* case, filing under seal permits the government to investigate the fraud and determine whether to prosecute without tipping off the alleged wrongdoer. If the government decides to prosecute, the case is unsealed and proceeds as a matter of public record. Likewise, insurance regulators may file for conservatorship under seal. This action would entail securing court orders to stabilize an insurance company without disclosing to the public (or perhaps any wrongdoers who might seek to flee a sinking ship) that the insurance company requires regulatory intervention.¹³ More prosaically, certain types of eviction actions can be sealed. An example would include evictions against blameless tenants where the landlord was foreclosed upon.¹⁴

¹⁰705 ILCS 105/16(6).

¹¹E.g., *In re Sealed Case*. (1997). 121 F.3d 729.

¹²31 U.S.C. § 3730; 740 ILCS 175/1.

¹³215 ILCS 5/188.1(4).

¹⁴735 ILCS 5/9-121; *Habitat Co., LLC v. Peebles*. (2018). 2018 IL App (1st) 171420.

Wholly sealed proceedings are more common in criminal court than in civil court. The most frequently sealed criminal proceedings are likely grand jury proceedings, a famously confidential and unique aspect of the US legal system (Reagan & Cort, 2009, p. 4; Roesch, Zapf, & Hart, 2009, p. 182; Foster, 2019).¹⁵ Technically speaking, grand jury proceedings are not an aspect of the courts, but rather part of a prosecution. However, given the outward similarities, they merit at least a brief mention here (Ardia, 2017, pp. 1435–1436). Other types of cases in the criminal sphere that may be sealed include applications by law enforcement for certain types of warrants, so as to not tip off the target of the warrant, as well as domestic violence cases and associated temporary restraining orders, to protect the privacy of victims of abuse (Administrative Office of the Illinois Courts, 2022b, pp. 64–66).

Finally, aside from cases that might be sealed while they are pending, criminal cases are unique in that they may be subject to sealing *after* they have come to an end, as part of a policy decision to aid defendants' rehabilitation. Almost every US state permits sealing closed criminal cases (Wakefield, Bialous, & Apollonio, 2023, p. 2). This kind of sealing has expanded substantially in recent years with concerted efforts to decriminalize marijuana use and possession, and though relatively few persons eligible for relief take advantage of it, the number of closed criminal cases that are eligible for some form of seal is substantial (Wakefield et al., 2023, p. 7). In Illinois alone, estimates suggest more than 2.2 million people—seventeen percent of the entire population—are eligible to have some portion of a criminal case sealed (Chien, Shaghagi, Madrigal, Thapa, & Gujarathi, 2023, p. 3).

3.2.2. Categorically Sealed Court Filings

In cases not already under a total seal, certain types of courtroom filings may still be removed from public access. Some documents are sealed in this way because of the type of document they are, with documents containing more personal or sensitive information more likely to be sealed. For example, pre-sentence investigations in criminal cases, which contain detailed information about a defendant's personal and family history, may be categorically sealed, even when the criminal case itself is otherwise open to the public (Ardia, 2017, p. 1436). More frequently, though, these categorical exclusions are based on specific information in a document. For example, if a civil case requires examining a person's criminal history, the criminal history might be filed under seal so as to maintain its confidentiality.¹⁶

Usually, categorical exclusions will be grounded in a statutory provision governing the sealing or other confidentiality requirements of a specific class of document (Ardia, 2017, pp. 1436–1438). Sealing in this way is intended to be limited in scope, and the remainder of the public record in the case should provide sufficient information to determine what was sealed and why (Reagan, 2010, p. 21).

3.2.3. Other Sealed Information

Theoretically speaking, any document in any case can be sealed for a variety of reasons. Parties may request, for example, that a given document be sealed to protect medical and business information (Reagan & Cort, 2009, p. 4); to prevent the public release of

¹⁵Fed. R. Crim. P. 6.

¹⁶28 C.F.R. § 23.2(f); 20 ILCS 2630/7.

sensitive documents such as psychological reports¹⁷ or search warrants;¹⁸ or simply because of the high profile of the parties involved (Reagan & Cort, 2009, p. 12).¹⁹ Whether these requests are granted, however, is another matter. Without a statutory basis for such a request, this type of sealing is both extraordinarily rare and highly disfavored by courts.²⁰

Courtroom filings may also be sealed as a matter of administrative policy. For example, if a document that inadvertently contains a complete bank account number is filed, one solution might be to file a new, redacted copy of that same document and seal the previously filed document. This strategy is particularly common in court filing systems that lack the technical capability to renumber or substitute filings. In such situations, a sealed court filing is not intended to be sealed as such, but rather the seal serves the mechanical purpose of correcting an error and withholding a document never intended to be public in the first place (Reagan & Cort, 2009, pp. 14–16).

3.3. Accessing Court Filings

All courtroom filings in the United States that have not been sealed are available for public inspection and copying at the clerk's office for the courthouse in which they were filed (Jones, 2013, p. 383). Practically speaking, however, physical access is rarely convenient and often not logistically possible. This section will therefore focus on methods of online records access (for more on the practical difficulties of acquiring records, see Ardia, 2017, pp. 1396–97).

3.3.1. Accessing Federal Court Documents

Federal courts started migrating to a digital platform in the 1990s (Martin, 2008, pp. 860–861); and since approximately 2007, all electronically stored documents in federal cases have been accessible through a records system known as “PACER” (Martin, 2008, p.864). This acronym stands for “Public Access to Court Electronic Records”. PACER accesses a database known by the admittedly less-catchy name of “Case Management/Electronic Case Files” or simply “CM/ECF”. Individual court filings are identified via the name of the court in which they were filed, the number of the case to which they belong, and the individual docket number of that filing. Anyone can create a PACER account to access the entire federal docket system and download individual filings as .pdf files.

PACER is a powerful research tool, but it has a significant potential drawback: cost. While access to judicial opinions is free, users are charged \$0.10 per page to access any court filing or case docket (United States Courts, 2019). These fees are waived, however, if a user accrues less than \$30 in a quarterly billing cycle. This “free” 300-page limit may not be sufficient for many researchers, however. Those in need of access to substantial amounts of federal court documents for research purposes, fortunately, can apply for a waiver of PACER fees (PACER, n.d.). The imposition of access fees for PACER has long been controversial (Shultze, 2018) and there have been a number of legislative attempts to remove them. However, for the time being, the cost remains.

¹⁷ *In re Marriage of Nadwodney*. (2012). 2012 IL App (1st) 111130-U.

¹⁸ *People v. Harris*. (2011). 2011 IL App (4th) 100275, ¶¶36–38.

¹⁹ *People v. Kelly*. (2009). 397 Ill. App. 3d 232.

²⁰ *Khan v. Gramercy Advisors, LLC*. (2016). 2016 IL App (4th) 150435 (Steigmann, J., specially concurring).

Readers may be familiar with the US federal Freedom of Information Act, or “FOIA”, which permits private individuals to request government public records. It is worth mentioning here because, while FOIA is a powerful tool to obtain documents relating to the function of government (Regalia, 2015), it is explicitly not applicable to federal courts.²¹ State FOIA statutes will also exclude courts, as is the case in Illinois.²² FOIA requests may be useful for obtaining courtroom data in some highly specific situations. For example, if a government agency is involved in litigation, the agency may be subject to a FOIA request for filings in its possession. However, going through another source will almost always be both easier and faster than requesting documents via a FOIA request.

3.3.2. Accessing State Court Documents

While the fundamental principle of public access to courtroom documents remains true in all fifty states, each state has its own approach to online access to those documents. As a result, the process of accessing courtroom data in each state system will differ, and there may even be marked differences in the way to access documents from different courts in the same state (Martin, 2008). Compounding this difficulty, not all state courts have fully digitized their records. This means that in some cases, online access to those records is outright impossible (Baldwin, Eassey, & Brooke, 2020, p. 750). While there has been improvement on this front in recent years (Fabri, 2021), state courts demonstrate considerable variation in their digitalization progress. Given the changing nature of this particular legal landscape, this section will discuss the sorts of features that characterize various state court electronic records systems in broader terms.

Almost all state courts offer a way to access docketing information online. This includes the ability to search for a case and access basic information, such as when the case was filed and when its next court date may be. The docketing information usually identifies each court filing in a case and may provide a brief summary of those filings, though these summaries may only consist of a few words. These more basic systems normally do not include access to the filed documents themselves.

Fortunately, many state courts do go further than that and maintain public access portals that offer more than just barebones docket information. The larger the court system, the more comprehensive its system is likely to be. Some state systems are functionally equivalent to PACER inasmuch as they provide a way for the public to access the same information available to court staff. Such portals may offer direct access to downloadable court filings, or at least a way for a user to request access to specific filings.

As with PACER, some court systems charge a fee for document access. If a court has a well-developed portal, this might be akin to a typical online shopping experience in which court filings are selected, the user pays at checkout, and the documents are then sent. In smaller systems, this process may instead involve emailing the court clerk or providing a credit card number over the phone. State courts may or may not have academic fee waivers. Once again, this appears to largely be a factor of how large and well-developed the court records system is and how frequently researchers have contacted a given clerk’s office in the past.

²¹5 U.S.C. § 551(1)(B).

²²*Copley Press v. Administrative Office of the Courts*. (1995). 271 Ill. App. 3d 548, 553.

In some cases, online access may be limited based on the status of the user. Attorneys may have more access to electronic records than members of the general public, for example, or parties to a case might have unlimited access to records from that case while nonparties are only able to access related judicial opinions. A user's status may also affect whether fees are assessed. For example, attorneys might have free access to records that would otherwise be subject to an access fee.

As a practical matter, and based on the personal experience of the authors, regardless of how any specific court system may be set up, one should never underestimate the utility of a polite phone call or email when attempting to acquire data that is meant to be publicly available. Clerk's offices exist to maintain records for the public, after all. They are therefore the best source for information about how their offices are managed as well as advice on how their records may be most easily obtained.

3.3.3. Third-Party Sources of Courtroom Data

The practical difficulties of accessing courtroom data have resulted in a number of third-party sources that attempt to provide easier access to these materials. Because the filings themselves are available to the public without restriction, third parties may host and redistribute them at no charge and without condition. Perhaps the best-known of these sources is Project RECAP, a free public database of millions of PACER documents overseen by the non-profit Free Law Project (Free Law Project, n.d.). There are many other such databases, though the smaller they are, the more they tend to specialize in certain types of documents. Justia, for example, is another very well-regarded site that focuses on providing judicial opinions and produces basic legal research guides (Justia, n.d.).

Most third-party databases focus on federal documents because, as was discussed above, accessing state court records can often be more difficult. Larger databases may include state court filings alongside federal filings, but they are rarely an area of focus. Paid databases might also maintain some level of access to electronic courtroom filings. The two major American legal research services, Westlaw and LexisNexis, each offer paid access to PACER and certain state filings, depending on the subscriber's level of access.

3.4. Using American Courtroom Data

So long as it does not fall into one of the sealable categories discussed above, the act of filing a document in court makes it a public record, and therefore accessible to the public, so long as that document has some connection to the case or the judicial process in which it was filed.²³ This presumption of public access applies equally to all people regardless of citizenship. This policy means that foreign nationals are entitled to exactly the same type of access to American courtroom data as US citizens are. What's more, once a given piece of courtroom data has been acquired by an individual, it can generally be used without restriction. The remainder of this section examines what this public record status means at a practical level.

²³ *United States v. Amodeo*. (1995). 44 F.3d 141, 146.

3.4.1. Courtroom Filings

As a general rule, courtroom filings may be used, analyzed, and distributed without restriction or any form of “continuing oversight” (Ardia, 2017, pp. 1413–1414). A leading and illustrative US Supreme Court case on this point is *Cox Broadcasting Corp. v. Cohn*.²⁴ In that case, a young woman was brutally sexually assaulted and murdered.²⁵ A reporter examined court files, learned her name, and broadcast a news report naming her.²⁶ The woman’s father filed suit. The legal basis of his suit was a Georgia state law that criminalized the publication of a rape victim’s identity.²⁷ The US Supreme Court held that the broadcast was permissible because it was derived from information contained in publicly available documents: “Once true information is disclosed in public court documents open to public inspection, the press cannot be sanctioned for publishing it”.²⁸ The Court noted that the appropriate way to protect the information would have been to ensure that it had not been publicly available in the first place, for example by sealing the proceedings, in whole or in part. Given the fact that such precautions had not been taken, once the woman’s information was made available in a public record, its dissemination could no longer be legally proscribed.²⁹

Though few public records cases involve such sensitive issues, the application of this principle in such an extreme circumstance makes clear: once a record is made available, it may not be subject to further restriction. Of course, courtroom filings may not be used to harass the subject of those filings.³⁰ But that harassment is impermissible regardless of the method chosen. In other words, though information obtained from courtroom filings may at times be used in the course of illegal activity, the use of those filings will not be the reason that the activity in question is illegal.

It is important to clarify here that if a document is already protected or its use restricted before filing, the fact of its filing likely does not strip those preexisting protections (Whiteman, 2005, p. 478).³¹ For example, the parties involved in a copyright lawsuit might file documents previously protected by copyright. The fact that a copyrighted document was filed in court does not strip its protections, and an individual who then sold a copy of a publicly filed copyrighted document would still be liable for copyright infringement (otherwise, no one would ever file such a lawsuit!). Fortunately for researchers, to the extent that any filing may be subject to copyright or similar protections, it would equally be subject to exceptions like those outlined by the fair use doctrine.³² Generally, if the document could be used for research purposes if obtained outside the court system, it remains usable if obtained as a courtroom filing.

3.4.2. Recordings of Courtroom Proceedings

The increasing prevalence of remote court proceedings has also led to a concomitant increase in the number of recordings of those proceedings. How (or even if) those

²⁴ *Cox Broadcasting Corp. v. Cohn*. (1975). 420 U.S. 469.

²⁵ *Id.* at 471.

²⁶ *Id.* at 472–74.

²⁷ *Id.* at 474.

²⁸ *Id.* at 496.

²⁹ *Id.* at 496.

³⁰ *Nixon v. Warner Comm’s.*(1978). 435 U.S. 589, 598 (collecting cases).

³¹ *Cox Broadcasting Corp. v. Cohn*. (1975). 420 U.S. 469, 496.

³² 17 U.S.C. § 107.

recordings can be used, however, varies from jurisdiction to jurisdiction. If a recording is obtained from an official source, it may be used without restriction. If, however, the origin of a recording is in doubt, the researcher should exercise caution.

Presumptively, “a trial is a public event”, and court proceedings are therefore open to the public.³³ But the ability to attend a court proceeding is not the same as the ability to record a court proceeding.³⁴ Many courts prohibit recordings of any sort, on penalty of contempt of court. Illinois, for example, explicitly bars recording of any proceedings, in-person or remote, in any form, unless specifically authorized by the judge in that case.³⁵ The United States Supreme Court famously bars all photography and video recording, though audio recordings and transcripts are released as a matter of course through its official website (Supreme Court of the United States, 2018).

Recordings of court proceedings are sometimes made by an official source such as a court’s website or a court-maintained YouTube channel. A famous example is the now viral “I Am Not a Cat” clip from a 2021 Texas proceeding (394th District Court of Texas – Live Stream, 2021). When such official recordings have been released, they too may be used without restriction. This is because the court itself is a government actor and therefore cannot claim copyright in the recording, and because third parties such as YouTube are similarly barred from claiming copyright in government works. If a recording has been released by a media organization, it is fair to assume that the recording was approved by the court and may likewise be used without restriction. With that said, while the courtroom content of such a report is non-copyrightable, any reportage surrounding it may still be subject to copyright, although any potential copyright claims would, once again, be subject to fair use exceptions. If a recording appears to have been made surreptitiously, further research is warranted. It would be worth determining, for example, whether the relevant court’s local rules bar recordings. If doubt remains, it is advisable to contact a judge’s chambers directly to obtain the clearest answer.

4. “Legal” Does Not Necessarily Mean “Ethical”

Just because the acquisition and use of American courtroom data for research purposes is legal does not mean that it is automatically ethical. As is true in all research, there are other factors that must be considered. It is not possible for this article to provide a comprehensive review of the many potentially relevant ethical considerations involved in research using American courtroom data, but there is one factor that should be considered in all such research: that there are real people involved in real disputes behind these texts. Very few litigants in the United States are likely aware that the documents created as part of their legal disputes could someday be studied by strangers and their personal information (or at least portions of it) used and distributed without their knowledge or consent. This does not mean that it is necessary to acquire the consent of the litigants named in a piece of courtroom data before examining it. In fact, the authors of this article do not consider obtaining such consent to be a necessary or even desirable step in research involving American courtroom data. While it may seem like a good idea in theory, there will be many circumstances in which contact is either impossible

³³*Craig v. Harney*. (1947). 331 U.S. 367, 374.

³⁴See *Westmoreland v. CBS*. (1984). 752 F.2d 16, 23–24.

³⁵E.g. Ill. Sup. Ct. R. 44.

for practical reasons or otherwise inadvisable (see, e.g., the discussion of whether to contact potentially dangerous litigants in Griffin, 2022, pp. 92-93). It is all the more important, therefore, to handle American courtroom data with great care, and to give particular thought to the presence of personally identifying information in such data when publishing any results. The following section will discuss some of the types of personally identifying information commonly found in American courtroom data and make recommendations as to how to responsibly handle that information.

4.1. Personally Identifying Information in American Courtroom Data

As a general rule, courtroom filings begin with a caption, which is “the introductory part of a court paper stating the names of the parties, the name of the court, the docket or file number, and a description of the paper” (Garner, 2024b). Other information may be present in a caption as well, such as the address of a property whose ownership is being disputed in that case. A caption in a hypothetical civil action involving a piece of property would look something like the following:

**IN THE CIRCUIT COURT OF COOK COUNTY, ILLINOIS
COUNTY DEPARTMENT, CHANCERY DIVISION**

David Griffin)	
PLAINTIFF)	No. 23 CH 12345
Vs.)	
Alex S. Moe)	1060 W. Addison St.,
DEFENDANT)	Chicago, IL 60613
)	

Figure 1. Legal caption from a hypothetical civil action involving property

The information contained in a caption such as Figure 1 can be very useful to a researcher. As was mentioned above, a case number, for example, can be used to identify other documents from the same case in records systems like PACER. Once that initial utility has been exhausted, however, a researcher should consider whether that personally identifying information has any ongoing use. Sometimes, as in cases involving public figures, the presence of such information will be a focal point of the research. Where this information is not so fundamentally relevant, however, a researcher should consider obscuring or anonymizing the personally identifying information of the parties involved. As explained earlier, because courtroom data is in the public domain, such precautions are not, strictly speaking, a legal requirement. However, from an ethical standpoint, it is always best to err on the side of caution, particularly in cases where the data examined in a study will later be made available alongside the study itself.

Given the fact-intensive nature of legal proceedings, many courtroom filings will contain personally identifying information (e.g., a reference to a person’s workplace or a relevant medical condition) throughout, not just in their caption. In some cases, this information will have been redacted prior to a given document being made publicly

available,³⁶ but there is no guarantee of this having happened. A close reading of the full text of each courtroom filing being examined is therefore recommended to identify what personally identifying information a given text contains and, where relevant, the best method with which it may be obscured or anonymized. That method, of course, will ultimately be dependent upon the data needs of a given project.

4.2. Data Storage

Even for studies in which all the personally identifying information in the data examined is going to be fully anonymized, there will be a period of time between the commencement of data collection and the completion of that anonymization process in which a researcher is in possession of courtroom documents that potentially contain sensitive information. It is therefore worth considering from the inception of such a project the way that data will be stored. This article is not able to make specific recommendations as to the best way to store potentially sensitive data in either physical or electronic formats; not only does the constantly changing landscape of available software and relevant legal regulations (e.g. the European Union's General Data Protection Regulation) mean any such recommendation would be almost instantly out of date, but it is also the experience of the authors that academic institutions and private companies that handle such data often have internal best practices that researchers will be obligated to follow. When in doubt, the best course of action is to either check with the relevant IT department or to consult the most appropriate up to date guidelines for ethical practices in applied linguistics (e.g. American Association for Applied Linguistics, 2025; British Association for Applied Linguistics, 2021).

5. Conclusion

There are very few legal restrictions on the acquisition and use of American courtroom data. It is hoped that this article has provided a clear guide on what those restrictions are and how a researcher can go about navigating them. At the same time, it is worth repeating that just because research in this area is legal (or perhaps better put, just because it is not illegal) does not mean that such research is necessarily ethical; that depends upon the nature of the research being carried out and the sensitivity of the researcher towards the very real people whose lives they are putting under a microscope. So long as that is kept in mind, however, and any personally identifying information is treated sensitively, there remains much to be gained from the continued study of this data.

Acknowledgements

The authors would like to thank Dr. Iman Nick and Dr. Kirsty Blewitt for their assistance in the preparation of an early draft of this manuscript.

³⁶Fed. R. Civ. P. 5.2; Fed. R. Crim. P. 49.1.

References

- 394th District Court of Texas – Live Stream. (2021, February). *Kitten Zoom Filter Mishap*. Retrieved from <https://www.youtube.com/watch?v=KxIPGPupdd8>
- Administrative Office of the Illinois Courts. (2022a). *Illinois Courts 2022 Annual Report* (Tech. Rep.). Retrieved from <https://ilcourtsaudio.blob.core.windows.net/antilles-resources/resources/d29322dd-c4d4-44ef-937b-a3761816f785/2022%20Annual%20Report%20Administrative%20Summary.pdf>
- Administrative Office of the Illinois Courts. (2022b). *Manual on Recordkeeping*. Retrieved from <https://ilcourtsaudio.blob.core.windows.net/antilles-resources/resources/89f29131-2311-4aa8-aa7f-2032921bc80a/Manual%20on%20Recordkeeping.pdf>
- American Association for Applied Linguistics. (2025). *AAAL Ethics Guidelines*. Retrieved from https://assets.noviams.com/novi-file-uploads/aaal/PDFs/AAAL_Ethics_Guidelines_-_App-a11b0518.pdf
- Ardia, D. S. (2017). Privacy and Court Records: Online Access and the Loss of Practical Obscurity. *University of Illinois Law Review*, 5, 1385–1454.
- Ardia, D. S., & Klinefelter, A. (2015). Privacy and Court Records: An Empirical Study. *Berkeley Technology Law Journal*, 30(2), 1807–1898.
- Baldwin, J. M., Eassey, J. M., & Brooke, E. J. (2020). Court Operations during the COVID-19 Pandemic. *American Journal of Criminal Justice*, 45, 743–758.
- British Association for Applied Linguistics. (2021). *Recommendations on Good Practice in Applied Linguistics* (4th Edition ed.). Retrieved from <https://www.baal.org.uk/wp-content/uploads/2021/03/BAAL-Good-Practice-Guidelines-2021.pdf>
- Chien, C. V., Shaghagi, N., Madrigal, A., Thapa, N., & Gujarathi, V. (2023). The Illinois Second Chance Expungement/Sealing Gap. *SSRN Electronic Journal*. Retrieved 2025-06-16, from <https://www.ssrn.com/abstract=4406408> doi: 10.2139/ssrn.4406408
- "The Common Law Right to Inspect and Copy Judicial Records: In Camera or On Camera". (1982). *Georgia Law Review*, 16(3), 659–693.
- Fabri, M. (2021, May). Will COVID-19 Accelerate Implementation of ICT in Courts? *International Journal for Court Administration*, 12(2), 2. Retrieved 2025-06-16, from <https://www.iacajournal.org/articles/10.36745/ijca.384/> doi: 10.36745/ijca.384
- Foster, M. A. (2019). *Federal Grand Jury Secrecy: Legal Principles and Implications for Congressional Oversight*. Congressional Research Service. Retrieved from <https://sgp.fas.org/crs/secrecy/R45456.pdf>
- Free Law Project. (n.d.). *Advanced RECAP Search*. Retrieved from <https://www.courtlistener.com/recap/>
- Garner, B. A. (Ed.). (2024a). *Black's Law Dictionary* (12th edition ed.). St. Paul, Minnesota: West Group.
- Garner, B. A. (Ed.). (2024b). *Caption (definition)* (12th edition ed.). St. Paul, Minnesota: West Group.
- Grainger, I., & Fealy, M. (1999). *An introduction to the new civil procedure rules*. London: Cavendish Publishing.
- Griffin, D. (2022). *Lexomancy: Law and Magic in the Pseudolegal Writings of the Sovereign Citizen Movement* (PhD thesis, Cardiff University, Cardiff). Retrieved from <https://orca.cardiff.ac.uk/id/eprint/155076>
- Incorporated Council of Law Reporting for England and Wales. (2022). *Plaintiff*. Re-

- rieved from <https://www.iclr.co.uk/knowledge/glossary/plaintiff/>
- Jones, D. R. (2013). Protecting the Treasure: An Assessment of State Court Rules and Policies for Access to Online Civil Court Records. *Drake Law Review*, 61, 375–422.
- Justia. (n.d.). *Justia*. Retrieved from <https://www.justia.com>
- Lombard, D. (2006). Top Secret: A Constitutional Look at the Procedural Problems Inherent in Sealing Civil Court Documents. *DePaul Law Review*, 55(3), 1067–1101.
- Martin, P. W. (2008). Online Access to Court Records - From Documents to Data, Particulars to Patterns. *Villanova Law Review*, 53, 855–887.
- PACER. (n.d.). *Fee Exemption Request for Researchers*. Retrieved from <https://pacer.uscourts.gov/my-account-billing/billing/fee-exemption-request-researchers>
- Reagan, T. (2010). *Sealing Court Records and Proceedings: A Pocket Guide*. Federal Judicial Center. Retrieved from https://northerndistrictpracticeprogram.org/wp-content/uploads/2017/05/Sealing_Guide.pdf
- Reagan, T., & Cort, G. (2009). *Sealed Cases in Federal Courts*. Federal Judicial Center. Retrieved from <https://www.uscourts.gov/sites/default/files/sealed-cases.pdf>
- Regalia, J. (2015). The Common Law Right to Information. *Richmond Journal of Law and the Public Interest*, 18(2), 89–132.
- Roesch, R., Zapf, P. A., & Hart, S. D. (2009). *Forensic Psychology and Law* (1st edition ed.). Wiley.
- Shultze, S. J. (2018). The Price of Ignorance: The Constitutional Cost of Fees for Access to Electronic Public Court Records. *The Georgetown Law Journal*, 106, 1197–1227.
- Supreme Court of the United States. (2018). *Transcripts and Recordings of Oral Arguments*. Retrieved from https://www.supremecourt.gov/oral_arguments/availabilityoforalargumenttranscripts.aspx
- United States Courts. (2019). *Electronic Public Access Fee Schedule*. Retrieved from <https://www.uscourts.gov/services-forms/fees/electronic-public-access-fee-schedule>
- Wakefield, T., Bialous, S., & Apollonio, D. E. (2023, April). Clearing cannabis criminal records: A survey of criminal record expungement availability and accessibility among US States and Washington DC that decriminalized or legalized cannabis. *International Journal of Drug Policy*, 114, 103983. Retrieved 2025-06-16, from <https://linkinghub.elsevier.com/retrieve/pii/S0955395923000324> doi: 10.1016/j.drugpo.2023.103983
- Whiteman, M. (2005). Appellate Court Briefs on the Web: Electronic Dynamis or Legal Quagmire? *Law Library Journal*, 97(3), 467–491.