Publication of court records: circumventing the privacy-transparency trade-off

Tristan Allard, Louis Béziaud, Sébastien Gambs

Publication of court records: circumventing the privacy-transparency trade-off*

Tristan Allard[1][0000−0002−2777−0027], Louis Béziaud[1,2][0000−0002−4974−3492], and Sébastien Gambs[2]

[1] Univ Rennes, CNRS, IRISA {tristan.allard,louis.beziaud}@irisa.fr
[2] Université du Québec à Montréal gambs.sebastien@uqam.ca

Abstract. The open data movement is leading to the massive publishing of court records online, increasing the transparency and accessibility of justice, and enabling the advent of legal technologies building on the wealth of legal data available. However, the sensitive nature of legal decisions also raises important privacy issues. Most of the current practices address the resulting privacy / transparency trade-off by combining access control with (manual or semi-manual) text redaction. In this work, we argue that current practices are insufficient for coping with the massive access to legal data, in the sense that restrictive access control policies are detrimental to both openness and to utility while text redaction is unable to provide sound privacy protection. Thus, we advocate for a integrative approach that could benefit from the latest developments in the privacy-preserving data publishing domain. We present a detailed analysis of the problem and of the current approaches, and propose a straw man multimodal architecture paving the way to a full-fledged privacy-preserving legal data publishing system.

Keywords: Privacy · Transparency · Legal data · Anonymization

1 Introduction

The opening of legal decisions to the public is one of the cornerstones of many modern democracies: it allows to audit and make accountable the legal system by ensuring that justice is rendered according to the laws in place. As stated in [9], it can even be considered that *"publicity is the very soul of justice"*. Additionally, in countries following the common law, the access to legal decisions is a necessity as the law in place emerged from the previous decisions of justice courts.

Thus, it is not surprising that the transparency of justice is enshrined in many countries as a fundamental principle, such as the *right to a public hearing* provided by the Article 6 of the European Convention on Human Rights, the Section 135(1) of the Courts of Justice Act (Ontario) stating the general principle that *"all court hearings shall be open to the public"* or in Vancouver

* A version of this work was presented at the Law and Machine Learning workshop at ICML 2020 (no proceeding).

Sun (Re) *"The open court principle has long been recognized as a cornerstone of the common law"*. The open data movement push for free access to law with for example the Declaration on Free Access to Law [16]. Multiple open government initiatives also consider the need for an open justice [49], such as the "Loi pour une République numérique" in France, the Open Government Partnership, the Open Data Charter and the Canada's Action Plan on Open Government.

Combined with recent advances in machine learning and natural language processing, the (massive) opening of legal data allows for new practices and applications, called legal technologies. Nonetheless, not all legal decisions should directly be published as such due to the privacy risks that might be incurred by victims, witnesses, members of the jury and judges. Privacy issues have been considered and mitigated by legal systems for a long time. For instance, the identities of the individuals involved in sensitive cases, such as cases with minors, are usually *anonymized* by default because they belong to a vulnerable subgroup of the population. In situations in which the risks of reprisal are high (*e.g.*, terrorism or organized crimes cases), judges, lawyers and witnesses might also ask for their identities to be hidden [21,26]. Finally, the identities of the members of a jury are also usually protected to guarantee that they will not be coerced but also to ensure that the strategy deployed by the lawyers is not tailored based on their background. Legal scholars are aware of the need for privacy when opening sensitive legal reports [8, 13, 25].

In the past, these privacy risks were limited due to the efforts required to access the decisions themselves. For instance, some countries require to go directly to the court itself to be able to access the legal decisions. Even when the information is available online, the access to legal decisions is usually on a one-to-one basis through a public but restricted API rather than enabling a direct download of the whole legal corpus. Typical restriction mechanisms include CAPTCHAs (SOQUIJ[3]), quotas (CanLII[4]), registration requirement as well as policy agreement and limitation of access to research scholars (Caselaw[5]). Furthermore, the fact that a legal decision is public does not mean that it can, legally, be copied and integrated in other systems or services without any restrictions.

A first approach to limit the privacy risks consists in *redacting* the legal decisions before publishing them. Redaction mostly follows predefined rules that list the information that must be removed or generalized and define how [48] (*e.g.*, by replacing the first and last names by initials, by a pseudonym). Redaction is in general semi-manual (and sometimes fully manual) because automatic redaction is error-prone [40]. This makes it extremely costly, not scalable and does not completely remove the risks of errors [48]. For example, 3.9 million decisions are pronounced in France every year but only 180000 are recorded in governmental databases and less than 15000 are made accessible to the public [22]. Moreover, even a perfect redaction would still offer weak privacy guarantees. A redacted text still contains a non-negligible amount of information, possibly identifying or

[3] `https://soquij.qc.ca`
[4] `https://www.canlii.org`
[5] `https://case.law`

sensitive, that may be extracted, *e.g.*, from the background of the case or even from the natural language semantics.

Another approach is access control, such as non-publication (*e.g.*, a case involving terrorism was held in secret in Britain [11]), rate limitation or registration requirements. However, access control mechanisms are binary and do not protect against privacy risks for the texts for which the access is granted. Furthermore, restricting massive accesses through blocking strategies also restricts the development of legal technologies that require a massive access to legal data.

In a nutshell, this paper makes the following contributions:

- We state the problem of reconciling transparency with privacy when opening legal data on a large scale (Section 2).
- We analyze the limits of the current approaches that are deployed in a widespread manner in real-life (Section 3).
- We propose a high-level straw man architecture of a system for publishing legal data massively in a privacy-preserving manner without precluding the traditional open court principles (Section 4).

2 Problem statement

Legal data. Legal reports are defined as written documents produced by a court about a particular judgment, which is itself a written decision of a court on a particular case. Although the content of a case report varies with courts and countries, it typically consist of elements such as date of hearing, names of judges and parties, facts, issue, etc. [59]

Need for readability and accessibility. The access to legal decisions is required both for transparency and practical reasons such as case law, which is the use of past legal decisions to support the decision for future cases. Thus, the judiciary system is built on the assumption that legal decisions are made public and accessible by default (*open-court principle*), so that (1) citizens are able to inspect decisions as a way to audit the legal system and (2) past decisions can be used to interpret laws, and as such must be known from legal practitioners and citizens. It follows that decisions must be made available in a form readable by humans (*i.e.*, natural language). The need for openness, the current practice in terms of open court, and the associated risks are detailed in [13, 41]. They conclude that, although there are powerful voices in favor of open court, radical changes in access and dissemination require new privacy constraints, and a public debate on the effect of sharing and using information in records.

Accessibility is also an important issue. In the past, the access to decisions required attending public hearings or reading books called "reporters". Today, web services share millions of decisions and facilitate access to legal records to individuals–law professionals (judges, lawmakers and lawyers), journalists, or citizens. Online publication also enables the large-scale access and processing of records, in particular due to a standardized format.

Need for massive accesses (legal technologies) The term *legal technologies* encompasses technologies used in the context of justice, such as practice management, analytics and online dispute resolution[6]. These applications often require some form of "understanding" of legal documents, usually performed through natural language processing (NLP) and machine learning (ML) approaches [15, 51]. We focus on this category as these applications are based on the analysis of a large number of legal data. One of the main challenges we have faced is that usually companies provide very few technical details about their actual processing and usage of legal documents.

The automatic processing and analysis of legal records have multiple applications, such as computing similarity between cases [38, 43, 58], predicting legal outcomes [3, 32] (*e.g.*, by weighing the strength of the defender arguments and the legal position of a client in a hypothetical or actual lawsuit), identifying influential cases [39, 45, 55] or important part of laws [44], estimating the risk of recidivism [57], summarizing legal documents [61], extracting entities (*e.g.*, parties, lawyers, law firms, judges, motions, orders, motion type, filer, order type, decision type and judge names) from legal documents [14, 52], topic modelling [6, 46], concept mapping [10] or inferring patterns [7, 35].

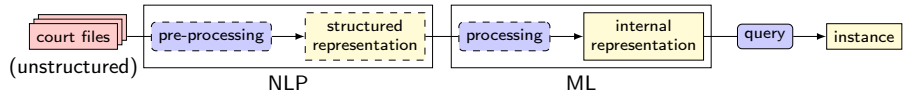

Fig. 1. High-level pipeline of court files processing for Legal Techs

Most of the technologies introduced in the previous section rely on the processing of large database of legal data. However, the unstructured nature of legal data is one of the main challenges of the application of artificial intelligence in law [2]. Consequently, the analysis of a legal text corpus first requires to apply some pre-processing to add structure to the text. Figure 1 represents an abstract processing pipeline for court files, extracted mostly from academic papers[7], and inferred from the current practice of text analysis and descriptions of associated technologies. In the following, we assume that any application involving the use of machine learning (as highlighted by most legal tech companies) is applied to court records. The first NLP step transforms the unstructured data (*i.e.*, natural language) into some structured representation (see below) by pre-processing it. Afterwards, the second ML step corresponds to the actual application, which is the training (*i.e.*, processing) of the ML algorithm, whose output is represented by the "internal representation" block. The term instance represents the output

[6] More examples are available at CodeX Techindex at `http://techindex.law.stanford.edu` which references more than a thousand companies.

[7] The majority of the legal technologies market consists in commercial applications. They do not give information about their inner working and underlying techniques.

of the model given some query (*e.g.*, applicable laws given a set of keywords representing infractions).

The pre-processing can be diverse and depends on the task (*e.g.*, extracting a citation graph between cases). However, most NLP-based applications usually rely on a text model. Many models are based on a bag-of-words (BoW) approach [27]. For example, document-word-frequency decomposes the text into a matrix in which each cell contains the number of times a particular word appears in a document. Other examples include term frequency-inverse document frequency and n-grams [63]. For example, a combination of those techniques are used in [3] to predict decisions from the European Court of Human Rights, and by [33] to identify law articles given a query or to answer to questions given a law article. Another common approach is word embeddings where words are mapped–using e.g. prediction-based or count-based methods–to real-valued vectors along with the context in which they are used [42]. Multiple variations of this structure exist [29, 34, 36, 37, 64]. This approach has been used for example in [39] to rank and explain influential aspects of law, or by [44] to predict the most relevant sources of law for any given piece of text using "neural networks and deep learning algorithms".

Need for privacy The massive opening of legal decisions for transparency and technological reasons must not hinder the right to privacy as emphasized by current open justice laws. In particular in this setting, the privacy of at least three main actors must be guaranteed: namely the individuals directly involved in decisions (*i.e.*, the parties), the individuals cited by decisions (*e.g.*, experts or witnesses), and the individuals administering the laws (*i.e.*, magistrates).

However, publishing legal decisions while providing sound privacy guarantees is difficult. For instance, authorship attacks [1] may lead to the re-identification of magistrates behind written decisions, or the presence of *quasi-identifiers*[8] within the text decisions may lead to the re-identification of the individuals involved or cited. Famous real-life examples, such as the governor Weld's [56] or Thelma Arnold's re-identification [5], both based on the exploitation of quasi-identifiers, are early demonstrations of the failure of naive privacy-preserving data publishing schemes. Thus despite the fact that legal decisions are written as unstructured text, structured information can be extracted from them, including the formal argument, the decision itself (*e.g.*, "guilty" or "innocent"), as well as arbitrary information about the individuals involved (*e.g.*, gender, age and social relationships).

Pseudonymization schemes simply consist in removing or replacing (e.g. by chainable or non-chainable pseudonyms) directly identifying data (*e.g.*, social security number, first name and last name, address) and keeping unchanged the rest of the information (quasi-identifiers included). These schemes provide a very weak protection level, as acknowledged by privacy legislations (*e.g.*, GDPR),

[8] A quasi-identifier is a combination of (one or more) attributes that are usually unique in the population, thus indirectly identifying an individual. A typical example is the triple `(age, zip code, gender)`.

which has led to the development of new approaches for sanitizing personal data in the last two decades (see for instance the survey in [12]). In this paper, we focus on privacy-preserving data publishing schemes providing formal privacy guarantees that hold against several publications (as required by any real-life privacy-preserving data publishing system). These schemes are based on (1) *a formal model* stating the privacy guarantees the scheme as well as one or more *privacy parameters* for tuning the "privacy level" that must be achieved, and (2) *a sanitization algorithm* designed to achieve the chosen model.

A formal model exhibits a set of *composability properties* that defines formally the impact on the overall privacy guarantees of using the scheme on a *log of publications* (also called *disclosures log* in the following). In particular, we will consider the ϵ-differential privacy model [17], defined formally in Definition 1, parametrized by ϵ, and achievable by the Laplace mechanism. Its self-composability properties are stated in Theorem 1 and its overall privacy guarantees are quantified by the evolution of the disclosures log, and in particular by the evolution of the ϵ value along the various differentially-private releases.

Definition 1 (ϵ-differential privacy [17]). *A randomized mechanism $\mathcal{M}$ satisfies ϵ-differential privacy, in which $\epsilon > 0$, if:*

$$\Pr[\mathcal{M}(\mathcal{D}_1) = \mathcal{O}] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(\mathcal{D}_2) = \mathcal{O}]$$

for any set $\mathcal{O} \in Range(\mathcal{M})$ and any tabular dataset $\mathcal{D}_1$ and $\mathcal{D}_2$ that differs in at most one row (in which each row corresponds to a distinct individual).

In a nutshell, ϵ-differential privacy ensures that the presence (or absence) of data of a single individual has a limited impact on the output of the computation, thus limiting the inference that can be done by an adversary about a particular individual based on the observed output.

Theorem 1 (Sequential and parallel Composability [19]). *Let $\mathtt{f}_i$ be a set of functions such that each provides ϵ_i-differential privacy. First, the* sequential composability *property of differential privacy states that computing all functions on the same dataset results in satisfying $(\sum_i \epsilon_i)$-differential privacy. Second, the* parallel composability *property states that computing each function on disjoint subsets provides $\mathtt{max}(\epsilon_i)$-differential privacy.*

3 Analysis of current practices

In the following section, we review the current practice for legal data anonymization and privacy regulations. To be concrete, we illustrate the privacy risks through examples of re-identification attacks. Finally, we argue that rule-based anonymization is not sufficient to provide a strong privacy protection and discuss the (formal) issues surrounding text anonymization.

3.1 Redaction *in the wild*

Redaction of legal data The redaction process consists in removing or generalizing a set of predefined terms defined by law through a semi-manual process [48]–e.g., using "find and replace" or domain-specific taxonomies combined with named entity recognition. Furthermore, access to legal documents or even public hearings can be restricted in well-defined cases. The common practice is to replace sensitive terms, as defined below, by initials, random letters, blanks or generalized terms (*e.g.*, "Montréal" becomes "Québec"). The specific set of rules regarding protected terms and the associated replacement practice can differ between countries and courthouses [48].

According to [50], information such as names, date and place of birth, contact details of unique identifiers (*e.g.*, social security number) is to be systematically removed for any person (subject to a restriction on publication), as well as for each of his or her relatives (*e.g.*, parents, children, neighbors, employers). In some contexts, additional information such as community or geographic location, intervenors (*e.g.*, court experts, social workers), or unusual information is also removed if it can be used to identify an individual. [13] presents numerous examples of legislation putting restriction to the *open-court principle*, such as hiding the identity of victims of sexual offenses.

Paper versus digital The main difference between paper and digital access is the "practical obscurity" of paper records on the one hand, and the easy accessibility of digital records, on the other. The awkwardness of accessing paper records stored in a public courthouse puts inherent limitations on the ability of individuals or groups to access those records. In contrast, digital records are easy to analyze, can be searched in "bulk" by combining various key factors (*e.g.*, divorce and children) and can potentially be accessed from any computer. Thus, traditional distribution provides "practical obscurity" [30], in that it is inconvenient (*i.e.*, time-consuming) to attend the courthouse or read case reports.

3.2 Limits of current approaches

In this section we provide examples of potential attacks in order to illustrate the technical difficulties of raw text anonymization. Figure 2 presents excerpts from French and Canadian opinions[9]. More examples are available in [4].

Figure 2a is anonymized according to the CNIL recommendations of 2006, which requires the last name of individuals to be replaced by its initial. However, widely available background knowledge on the "Real Madrid Club de Futbol" combined with the (real-life) pseudonyms of the "players" trivially leaks their identity.

The de-anonymization of Figure 2b relies on the text semantics instead of background knowledge. It requires the adversary (1) to identify the link (X) between "M. [...] Abdel X" and "the use of the name 'X' to designate a drink",

[9] We translated them using DeepL (`https://www.deepl.com`)

the association Real Madrid Club de Futbol and several players of this team, Zinedine Z., David B., Raul Gonzalès B. aka Raul, Ronaldo Luiz Nazario de L., aka Ronaldo, and Luis Filipe Madeira C., aka Luis Figo

(a) CA Paris, 14 févr. 2008, n° 06/11504

the American company Coca Cola Company markets drinks under the French trade mark "Coca Cola light sango", of which it is the proprietor; that M. [...] Abdel X, relying on the infringement of his artist's name and surname, has brought an action for damages against the Coca Cola Company [...] On the ground that Abdel X maintains that, as an author and screenwriter, he is entitled to oppose the use of the name "X" to designate a drink marketed by the companies of the Coca Cola group.

(b) Cass. 1re civ., 10 avr. 2013, n° 12-14.525, Bull. 2013, I, n° 72.

X, born [...] 2017; Y, born [...] 2018 the children and C; D the parents

Applications are submitted for X, aged 1 year, and Y, aged 2 months. The Director of Youth Protection (DYP) would like X to be entrusted to her aunt, Ms. E, until June 25, 2019. As for Y, that he be entrusted to a foster family for the next nine months. The father has two other children, Z and A, from his previous union with Mrs. F. The mother has another child, B, from her union with Mr. G.

(c) Protection de la jeunesse — 201518, 2020 QCCQ 10887

Fig. 2. Excerpts of legal decisions

and (2) to infer that the drink is called "sango", thus leading to the conclusion that X = "sango". While this attack may not be easy to automatize due to the hardness of detecting the semantics inference, it is, however, trivial to perform for a human (*e.g.*, by crowdsourcing it).

Figure 2c could be attacked through a combination of attributes and relationship. This opinion from the Youth court involves children and, as such, follows the strictest anonymization rules of the SOQUIJ. However, an adversary can extract an extensive relationship graph which could be matched over a relationship database (*e.g.*, Facebook).

Besides the content of legal documents, stylometry [47] can also be used to identify authors (*i.e.*, magistrates) by their writing style. Mitigation for this kind of attack exist [20, 60] but their output is only machine readable (i.e., they do not fulfill the readability requirement, but are of intereset when considering "massive" processing in Section 4). Similarly, it is possible to exploit decision patterns to re-identify judges, as done for the Supreme Court of the United States [32].

3.3 Reasons for the failure of rule-based redaction

Reviews of current practices for tackling the privacy of legal documents in Section 3.1 has highlighted the widespread use of rule-based redaction, in which a set of patterns is defined as being sensitive and is either removed or replaced. However, as shown in Section 3.2 (1) privacy can be violated even in "simple" instances and (2) identifying information remains in most cases. In other words, rule-based redaction does not provide any sound privacy guarantee.

We observe that it suffers from the following main difficulties. (1) *Missing rule*: many combinations of quasi-identifiers can lead to re-identification and

the richness of the output space offered by natural language (*i.e.*, what can be expressed) can hardly be constrained to a set of rules. (2) *Missing match*: The current state of the art about relationship extraction and named-entity recognition makes it hard to ensure that all terms that should be redacted will be detected, in particular because of the many possible ways to express the same idea (*e.g.*, *circumlocution*).

Although these observations make the rule-based redaction difficult, it is important to note that attacks, *e.g.*, re-identification, remain simpler than protection. Indeed, an adversary has to find a single attack vector (*i.e.*, a missing rule or a missing pattern) whereas the redaction process needs to consider all the possibilities.

4 Multimodal publication scheme

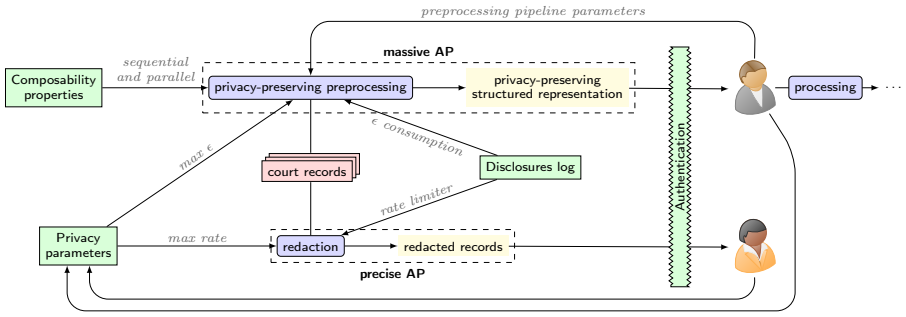

Fig. 3. Multimodal publication architecture

In Section 2, we have shown that the publication of legal documents serves two distinct and complementary purposes: (1) the traditional objective of transparency and case law, and (2) the modern objective of legal technologies of providing services to citizens and legal professionals. These two purposes obey to different utility and privacy requirements. More precisely, the traditional use case requires human-readable documents while legal techs need a machine-readable format for automated processing. Moreover, transparency and case law involve the access to opinions on an individual basis (*i.e.*, one-at-a-time), similarly to attending a hearing in person. In contrast, legal technologies rely on the access to massive legal databases. This difference in cardinality (*i.e.*, one versus many) entails different privacy risks. In particular, the massive processing of legal data requires the use of a formal privacy framework with composability properties (see Section 2). All this suggests the inadequacy of any *one-size-fits-all* approach. As a consequence, we propose that the organization in charge of the publication of

legal decisions consider two modes of publication[10]: the *precise access mode* and the *massive access mode*.

Precise access mode To fulfill the "traditional" use case, the precise access mode provides full access to legal decisions that are only redacted using the current practices. This access mode is designed for the transparency and case law usages, and is to be used typically by individuals (*e.g.*, law professionals, journalists and citizens). Similar to the "traditional" paper-based publication scheme, in the precise access mode [23], a user has access to full and partial documents. While the current practice of redacting identifiers could be combined with more automated approaches such as [24,54]. The aim of this mode is to provide strong utility first. It allows browsing, searching and reading documents similar to the websites currently publishing legal documents (*e.g.*, Legifrance or CanLII).

To prevent malicious users from diverting the precise access mode for performing massive accesses, users must be authenticated and their access must be restricted (*e.g.*, rate limitation or proof of work [18]). The access restrictions of a given user can be tuned depending on his trustworthiness (e.g., strength of the authentication, legally binding instruments implemented). The main objective of the restricted access mode is to make it difficult to rebuild the full (massive) database.

Massive access mode The massive access mode gives access only to pre-processed data resulting from privacy-preserving versions of the standard NLP pipelines available on the server, *i.e.*, aggregated and structured data extracted from or computed over large numbers of decisions, as required for the "modern" use case. It should be compatible with most legal tech applications that traditionally use a database of legal documents (see Section 2). Note that the perturbations due to privacy-preserving data publishing schemes have usually less impact (in terms of information loss) when applied after aggregation (i.e., late in the pipeline, see Figure 3 or [53, Figure 1]), at the cost of a loss of generality of the output.

Users need to be able to tune the pre-processing applied. For the sake of simplicity, we assume that the user (*i.e.*, legal tech developer) provides the parameters for a given NLP pipeline (see Fig. 3). These parameters can be for instance the maximum number of features or n-grams range to consider. In order to avoid limiting the massive access mode to the current implementation state of its NLP libraries, more complex implementations can be considered (1) by generating synthetic *testing* data in a privacy-preserving manner (*e.g.*, PATE-GAN [28]) or (2) by relying on a full pre-processing pipeline that embeds privacy-preserving calls to the server (*e.g.*, through a privacy-preserving computation framework such as Ektelo [62]).

The massive access mode must also authenticate users in order to monitor the overall privacy guarantees satisfied for each user based on his disclosures log and on the composability properties of the privacy-preserving data publishing

[10] The technical protection measures can be strengthened by usual legal instruments (e.g., non-disclosure agreements).

schemes used. As a result, the data is protected using authentication and strong privacy definitions.

Finally, another potential need is the annotation of documents, which is the addition of metadata to terms, sentences, paragraphs or documents such as syntax (*e.g.*, verb), semantic or pragmatic (*e.g.*, implicature). This step is crucial in NLP, and is usually done manually, for example through crowdsourcing. Crowdsourcing-specific approaches for privacy-preserving task processing [31] require splitting the task (*i.e.*, annotation) between non-colluding workers before aggregating the result in a secure way (*e.g.*, on the platform).

System overview Figure 3 outlines an abstract architecture for our privacy-preserving data publishing system for legal decisions. Our objective is not to provide exhaustive implementation guidelines, but rather to identify the key components that such an architecture should possess. The precise and massive access modes are both protected by the `Authentication` module. The `Authentication` module can be implemented by usual strong authentication techniques (*e.g.*, for preventing impersonation attacks). Authentication is necessary for enforcing the access control policy through the `Access Control` module and for maintaining for each user his `Disclosure Log`. The log contains all the successful access requests performed by a user. It is required for verifying that the overall privacy guarantees are not breached, *e.g.*, the rate limitation is not exceeded, or the composition does not exceed the tolerated disclosure. Finally, the `Privacy Parameters` contain the overall privacy guarantees that must always hold, defined by the administrator (*e.g.*, rate limit or higher bound on the tolerated disclosure). The user may additionally be allowed to tune the privacy parameters input by a privacy-preserving data publishing scheme (*e.g.*, the fraction spent in the higher bound on the ϵ differential privacy parameter) provided it does not jeopardize the overall privacy guarantees.

5 Conclusion

In this paper, we analyzed the needs for publishing legal data and the limitations of rule-based redaction (*i.e.*, the current approach) for fulfilling them successfully. We proposed to discard any one-size-fits-all approach and outlined a straw man architecture balancing the utility and privacy requirements by distinguishing the traditional, one-to-one, use of legal data from the modern, massive, use of legal data by legal technologies. Our proposition can easily be implemented on current platforms.

Acknowledgments

We thank the reviewers for their careful reading of the manuscript and their constructive remarks. This work was partially funded by the PROFILE-INT project funded by the LabEx CominLabs (ANR-10-LABX-07-01). Sébastien Gambs is supported by the Canada Research Chair program, a Discovery Grant (NSERC) and the Legalia project (FQRNT).

References

1. Abbasi, A., Chen, H.: Writeprints: a stylometric approach to identity-level identification and similarity detection in cyberspace. ACM Transactions on Information Systems (TOIS) **26**(2), 7 (2008)
2. Alarie, B., Niblett, A., Yoon, A.H.: How artificial intelligence will affect the practice of law. University of Toronto Law Journal **68**(supplement 1), 106–124 (2018)
3. Aletras, N., Tsarapatsanis, D., Preoţiuc-Pietro, D., Lampos, V.: Predicting judicial decisions of the european court of human rights: a natural language processing perspective. PeerJ Computer Science (2016)
4. Allard, T., Béziaud, L., Gambs, S.: Online publication of court records: circumventing the privacy-transparency trade-off (2020)
5. Arrington, M.: AOL proudly releases massive amounts of private data. TechCrunch (2006), https://social.techcrunch.com/2006/08/06/aol-proudly-releases-massive-amounts-of-user-search-data/
6. Ashley, K.D., Brüninghaus, S.: Automatically classifying case texts and predicting outcomes. Artif. Intell. Law **17**(2), 125–165 (Jun 2009)
7. Ashley, K.D., Walker, V.R.: Toward constructing evidence-based legal arguments using legal decision documents and machine learning. In: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Law. pp. 176–180 (2013)
8. Bailey, J., Burkell, J.: Revisiting the open court principle in an era of online publication: questioning presumptive public access to parties' and witnesses' personal information. Ottawa L. Rev. **48**, 143 (2016)
9. Bentham, J., Bowring, J.: The works of Jeremy Bentham, vol. 4. William Tait (1843)
10. Brüninghaus, S., Ashley, K.D.: Using machine learning for assigning indices to textual cases. In: Leake, D.B., Plaza, E. (eds.) Case-Based Reasoning Research and Development. pp. 303–314. Springer Berlin Heidelberg, Berlin, Heidelberg (1997)
11. Calamur, K.: In a first for Britain, a secret trial for terrorism suspects. NPR (2014), https://text.npr.org/s.php?sId=319076959
12. Chen, B.C., Kifer, D., LeFevre, K., Machanavajjhala, A.: Privacy-preserving data publishing. Found. Trends Databases **2**(1–2), 1–167 (Jan 2009)
13. Conley, A., Datta, A., Nissenbaum, H., Sharma, D.: Sustaining privacy and open justice in the transition to online court records: a multidisciplinary inquiry. Md. L. Rev. **71**, 772 (2011)
14. Custis, T., Schilder, F., Vacek, T., McElvain, G., Alonso, H.M.: Westlaw edge AI features demo: KeyCite overruling risk, litigation analytics, and WestSearch plus. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law - ICAIL '19. pp. 256–257. ACM Press, Montreal, QC, Canada (2019)
15. Dale, R.: Law and word order: NLP in legal tech. Natural Language Engineering **25**(1), 211–217 (Jan 2019)
16. Declaration on free access to law (2002), http://www.worldlii.org/worldlii/declaration/
17. Dwork, C.: Differential privacy. In: Proceedings of the 33rd International Conference on Automata, Languages and Programming - Volume Part II. Icalp'06, vol. 4052, pp. 1–12. Springer-Verlag, Berlin, Heidelberg (Jul 2006)
18. Dwork, C., Naor, M.: Pricing via processing or combatting junk mail. In: Annual International Cryptology Conference. pp. 139–147. Springer (1992)

19. Dwork, C., Roth, A., et al.: The algorithmic foundations of differential privacy. Foundations and Trends® in Theoretical Computer Science **9**(3–4), 211–407 (2014)
20. Fernandes, N., Dras, M., McIver, A.: Generalised differential privacy for text document processing. In: International Conference on Principles of Security and Trust. pp. 123–148. Springer (2019)
21. Fleuriot, C.: Avec l'accès gratuit à toute la jurisprudence, des magistrats réclament l'anonymat. Dalloz Actualité (Feb 2017), https://www.dalloz-actualite.fr/flash/avec-l-acces-gratuit-toute-jurisprudence-des-magistrats-reclament-l-anonymat
22. Fouret, A., Perez, M., Barrière, V., Rottier, E., Buat-Ménard, É.: Open Justice. Tech. rep., Direction interministérielle du numérique (2019), https://entrepreneur-interet-general.etalab.gouv.fr/defis/2019/openjustice.html
23. Hartzog, W., Stutzman, F.: The case for online obscurity. Calif. L. Rev. **101**, 1 (2013)
24. Hassan, F., Sánchez, D., Soria-Comas, J., Domingo-Ferrer, J.: Automatic anonymization of textual documents: detecting sensitive information via word embeddings. In: 2019 18th IEEE International Conference On Trust, Security And Privacy In Computing And Communications/13th IEEE International Conference On Big Data Science And Engineering (TrustCom/BigDataSE). pp. 358–365 (2019)
25. Jaconelli, J.: Open justice: a critique of the public trial. Oxford University Press on Demand (2002)
26. Jacquin, J.B.: Terrorisme : la peur des magistrats. Le Monde (Jan 2017), https://www.lemonde.fr/police-justice/article/2017/01/19/terrorisme-la-peur-des-magistrats_5065242_1653578.html
27. Joachims, T.: Text categorization with support vector machines: learning with many relevant features. In: European conference on machine learning. pp. 137–142. Springer (1998)
28. Jordon, J., Yoon, J., van der Schaar, M.: PATE-GAN: generating synthetic data with differential privacy guarantees. In: 7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019. OpenReview.net (2019)
29. Joulin, A., Grave, E., Bojanowski, P., Mikolov, T.: Bag of tricks for efficient text classification. arXiv preprint arXiv:1607.01759 (2016)
30. Judges Technology Advisory Committee: Open courts, electronic access to court records, and privacy: discussion paper. Tech. rep., Canadian Judicial Council (2003), http://publications.gc.ca/collections/collection_2008/lcc-cdc/JL2-75-2003E.pdf
31. Kajino, H., Baba, Y., Kashima, H.: Instance-privacy preserving crowdsourcing. In: Second AAAI Conference on Human Computation and Crowdsourcing (2014)
32. Katz, D.M., Bommarito II, M.J., Blackman, J.: A general approach for predicting the behavior of the supreme court of the united states. PLoS One **12**(4) (2017)
33. Kim, M.Y., Rabelo, J., Goebel, R.: Statute law information retrieval and entailment. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law - ICAIL '19. pp. 283–289. ACM Press, Montreal, QC, Canada (2019)
34. Kim, Y.: Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882 (2014)
35. Kort, F.: Quantitative analysis of fact-patterns in cases and their impact on judicial decisions. Harv. L. Rev. **79**, 1595 (1965)

36. Lai, S., Xu, L., Liu, K., Zhao, J.: Recurrent convolutional neural networks for text classification. In: Twenty-ninth AAAI conference on artificial intelligence (2015)
37. Liu, P., Qiu, X., Huang, X.: Recurrent neural network for text classification with multi-task learning. arXiv preprint arXiv:1605.05101 (2016)
38. Mandal, A., Chaki, R., Saha, S., Ghosh, K., Pal, A., Ghosh, S.: Measuring similarity among legal court case documents. In: Proceedings of the 10th Annual ACM India Compute Conference. pp. 1–9. Compute '17, Association for Computing Machinery, New York, NY, USA (2017)
39. Marques, M.R., Bianco, T., Roodnejad, M., Baduel, T., Berrou, C.: Machine learning for explaining and ranking the most influential matters of law. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law. pp. 239–243. Acm (2019)
40. Marrero, M., Urbano, J., Sánchez-Cuadrado, S., Morato, J., Gómez-Berbís, J.M.: Named entity recognition: fallacies, challenges and opportunities. Computer Standards & Interfaces **35**(5), 482–489 (2013)
41. Martin, P.W.: Online access to court records-from documents to data, particulars to patterns. Vill. L. Rev. **53**, 855 (2008)
42. Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in neural information processing systems. pp. 3111–3119 (2013)
43. Minocha, A., Singh, N.: Legal document similarity using triples extracted from unstructured text. In: Rehm, G., Rodríguez-Doncel, V., Moreno-Schneider, J. (eds.) Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). European Language Resources Association (ELRA), Paris, France (May 2018)
44. Mokanov, I., Shane, D., Cerat, B.: Facts2Law: using deep learning to provide a legal qualification to a set of facts. In: Proceedings of the Seventeenth International Conference on Artificial Intelligence and Law. pp. 268–269. Acm (2019)
45. Možina, M., Žabkar, J., Bench-Capon, T., Bratko, I.: Argument based machine learning applied to law. Artificial Intelligence and Law **13**(1), 53–73 (2005)
46. Nallapati, R., Manning, C.D.: Legal docket classification: where machine learning stumbles. In: Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing. pp. 438–446 (2008)
47. Neal, T., Sundararajan, K., Fatima, A., Yan, Y., Xiang, Y., Woodard, D.: Surveying stylometry techniques and applications. ACM Computing Surveys (CSUR) **50**(6), 1–36 (2017)
48. Opijnen, M., Peruginelli, G., Kefali, E., Palmirani, M.: On-line publication of court decisions in the EU: report of the policy group of the project "building on the European case law identifier". Available at SSRN 3088495 (2017)
49. Organisation for Economic Co-operation and Development (ed.): The call for innovative and open government: an overview of country initiatives. OECD, Paris (2011)
50. Plamondon, L., Lapalme, G., Pelletier, F.: Anonymisation de décisions de justice. In: XIe Conférence sur le Traitement Automatique des Langues Naturelles (TALN 2004). pp. 367–376. Bernard Bel et Isabelle Martin. (éditeurs), Bernard Bel et Isabelle Martin. (éditeurs), Fès, Maroc (May 2004)
51. Praduroux, S., de Paiva, V., di Caro, L.: Legal tech start-ups: state of the art and trends. In: Proceedings of the Workshop on MIning and REasoning with Legal texts collocated at the 29th International Conference on Legal Knowledge and Information Systems (2016)

52. Quaresma, P., Gonçalves, T.: Using linguistic information and machine learning techniques to identify entities from juridical documents. In: Francesconi, E., Montemagni, S., Peters, W., Tiscornia, D. (eds.) Semantic Processing of Legal Texts: Where the Language of Law Meets the Law of Language, pp. 44–59. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
53. Rastogi, V., Hong, S., Suciu, D.: The boundary between privacy and utility in data publishing. In: VLDB (2007)
54. Sanchez, D., Batet, M., Viejo, A.: Automatic general-purpose sanitization of textual documents. IEEE Transactions on Information Forensics and Security **8**(6), 853–862 (2013)
55. Siegel, D.J.: CARA: an assistance to help find the cases you missed. Law Prac. **43**, 22 (2017)
56. Sweeney, L.: K-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. **10**(5), 557–570 (Oct 2002)
57. Tan, S., Adebayo, J., Inkpen, K., Kamar, E.: Investigating human+ machine complementarity for recidivism predictions. arXiv preprint arXiv:1808.09123 (2018)
58. Thenmozhi, D., Kannan, K., Aravindan, C.: A text similarity approach for precedence retrieval from legal documents. In: FIRE (Working Notes). pp. 90–91 (2017)
59. University of Houston Law Center: How to brief a case. Tech. rep., University of Houston Law Center (2009), `https://www.law.uh.edu/lss/casebrief.pdf`
60. Weggenmann, B., Kerschbaum, F.: Syntf: synthetic and differentially private term frequency vectors for privacy-preserving text mining. arXiv preprint arXiv:1805.00904 (2018)
61. Yousfi-Monod, M., Farzindar, A., Lapalme, G.: Supervised machine learning for summarizing legal documents. In: Farzindar, A., Kešelj, V. (eds.) Advances in Artificial Intelligence. pp. 51–62. Springer Berlin Heidelberg, Berlin, Heidelberg (2010)
62. Zhang, D., McKenna, R., Kotsogiannis, I., Hay, M., Machanavajjhala, A., Miklau, G.: EKTELO: a framework for defining differentially-private computations. In: Proceedings of the 2018 International Conference on Management of Data. pp. 115–130. SIGMOD '18, Association for Computing Machinery, New York, NY, USA (2018)
63. Zhang, X., Zhao, J., LeCun, Y.: Character-level convolutional networks for text classification. In: Advances in neural information processing systems. pp. 649–657 (2015)
64. Zheng, J., Guo, Y., Feng, C., Chen, H.: A hierarchical neural-network-based document representation approach for text classification. Mathematical Problems in Engineering **2018** (2018)