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

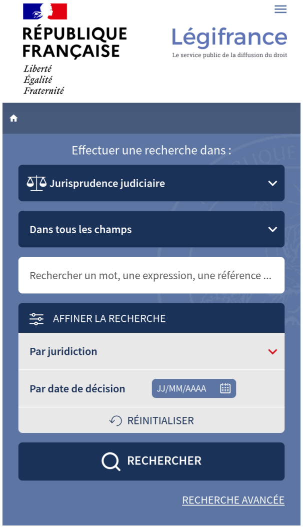
Tristan Allard, Louis Béziaud, Sébastien Gambs

IRISA, Univ Rennes 1, Université du Québec à Montréal

AICOL 2020

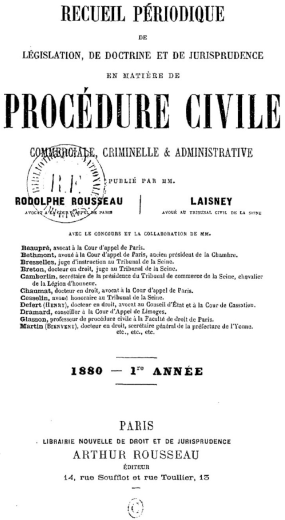
Big (legal) data

- Public online access to massive number of court records
  - 1 M sur Légifrance (France)
  - 2.7 M sur CanLII (Canada)
  - 6.7 M sur Caselaw (USA)
- Natural language + (few) meta-data



Open justice

- “Publicity is the very soul of justice” (Bentham)
- Transparency
  - ☞ Trust, bias inspection
- Accessibility
  - ☞ Utility constraint (case law)
- Paradigm shift
  - from paper-based and in-person cout hearings to electronic records
- Open-government projects (OECD, OGP, OGI)
- Massive processing
  - ☞ New technologies: Legaltechs



RECUEIL PÉRIODIQUE

DE

LÉGISLATION, DE DOCTRINE ET DE JURISPRUDENCE

EN MATIÈRE DE

PROCÉDURE CIVILE

COMMERCIALE, CRIMINELLE & ADMINISTRATIVE

PUBLIÉ PAR MM.

RODOLPHE ROUSSEAU
AVOCAT À LA COUR D'APPEL DE PARIS

LAISNEY
AVOUÉ AU TRIBUNAL CIVIL DE LA SEINE

AVEC LE CONCOURS ET LA COLLABORATION DE MM.

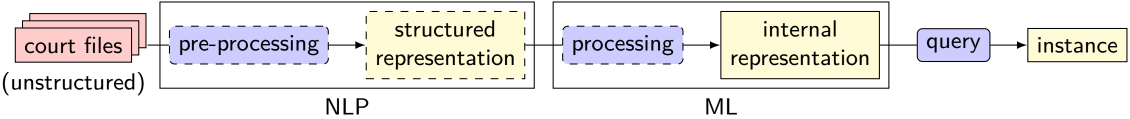
Beaupré, avocat à la Cour d'appel de Paris.
Bethmont, avoué à la Cour d'appel de Paris, ancien président de la Chambre.
Bresselles, juge d'instruction au Tribunal de la Seine.
Breton, docteur en droit, juge au Tribunal de la Seine.
Camberlin, secrétaire de la présidence du Tribunal de commerce de la Seine, chevalier de la Légion d'honneur.
Chaumat, docteur en droit, avocat à la Cour d'appel de Paris.
Cesselin, avoué honoraire au Tribunal de la Seine.
Defert (Henry), docteur en droit, avocat au Conseil d'État et à la Cour de Cassation.
Dramard, conseiller à la Cour d'Appel de Limoges.
Glasson, professeur de procédure civile à la Faculté de droit de Paris.
Martin (Bienvenu), docteur en droit, secrétaire général de la préfecture de l'Yonne.
etc., etc., etc.

1880 — 1re ANNÉE

PARIS
LIBRAIRIE NOUVELLE DE DROIT ET DE JURISPRUDENCE
ARTHUR ROUSSEAU
ÉDITEUR
14, rue Soufflot et rue Toullier, 13

Legal technologies

- Use court records for document automation, e-discovery, analytics, etc
- Fast expanding market[1]
- Buzzwords IA + NLP



[1] LawGeex. “Legal Tech Buyer’s Guide”. In: (2019).

Privacy risks

- Juges fear retaliation and coercion (mafia, terrorism, etc)[2][3]
- Risk of legal optimization (eg. judge analytics)
- “Google is linking secret, court-protected names–including victim IDs–to online coverage”[4]
- Linking to other databases for profiling, risk scoring, etc

[2] Jean-Baptiste Jacquin. “Terrorisme : la peur des magistrats”. In: *Le Monde* (Jan. 2017).

[3] Caroline Fleuriot. “Avec l’accès gratuit à toute la jurisprudence, des magistrats réclament l’anonymat”. In: *Dalloz Actualité* (Feb. 2017).

[4] Andrew Duffy. “Google is linking secret, court-protected names including victim IDs to online coverage”. In: *Ottawa Citizen* (2017).

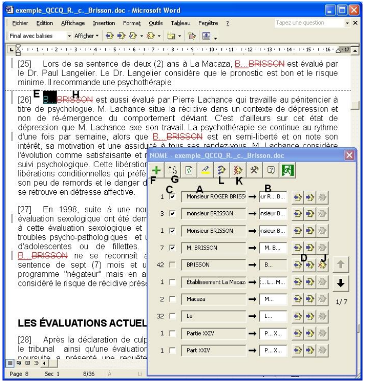
Current privacy protection

- Pseudonymization of names (with court-level rules)
- Manual process (with software support) in 75% of EU countries[5]
- Approach similar to medical data anonymization: “search and replace”
- Few research on anonymization of natural language with formal privacy guarantees

✓ Human readable

✓ Analytics

✗ Privacy



exemple_QCCQ_R._c._Brisson.doc - Microsoft Word

[25] Lors de sa sentence de deux (2) ans à La Macaza, B... ~~BRISSON~~ est évalué par le Dr. Paul Langelier. Le Dr. Langelier considère que le pronostic est bon et le risque minime. Il recommande une psychothérapie.

[26] B... ~~BRISSON~~ est aussi évalué par Pierre Lachance qui travaille au pénitencier à titre de psychologue. M. Lachance situe la récidive dans un contexte de dépression et non de ré-émergence du comportement déviant. C'est d'ailleurs sur cet état de dépression que M. Lachance axe son travail. La psychothérapie se continue au rythme d'une fois par semaine, alors que B... ~~BRISSON~~ est en semi-liberté et on note son intérêt, sa motivation et une assiduité à tous ses rendez-vous. M. Lachance considère l'évolution comme satisfaisante et r suivi psychologique. Cette libération libérations conditionnelles qui préfè son peu de remords et le danger d se retrouve en détresse affective.

[27] En 1998, suite à une nou évaluation sexologique ont été dem à cette évaluation sexologique et troubles psycho-pathologiques et u d'adolescentes ou de fillettes. B... ~~BRISSON~~ ne se reconnaît a sentence de sept (7) mois et u programme "négateur" mais en a considéré le risque de récidive prése

LES ÉVALUATIONS ACTUEL

[28] Après la déclaration de culp le tribunal ainsi qu'une évaluation poursuite a présenté une requête

NOME - exemple_QCCQ_R._c._Brisson.doc

1	☑	Monsieur ROGER BRISS	→	ur R... B...
3	☑	monsieur BRISSON	→	nsieur B...
1	☑	Monsieur BRISSON	→	nsieur B...
7	☑	M. BRISSON	→	M. B...
42	□	BRISSON	→	B...
1	□	Établissement La Macaz	→	É... L... M...
2	□	Macaza	→	M...
32	□	La	→	L...
1	□	Partie XXIV	→	P... X...
1	□	Part XXIV	→	P... X...

[5] Marc Opijnen et al. “On-line publication of court decisions in the EU”. In: (2017).

Redaction in practice

Cour d'appel de Paris 11ème chambre, section B Arrêt du 14 février 2008

> "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"

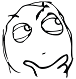
- follows the privacy recommendations of the CNIL from 2006

☞ background knowledge and pseudonyms

Redaction in practice

Protection de la jeunesse – 186470, 2018 QCCQ 6920

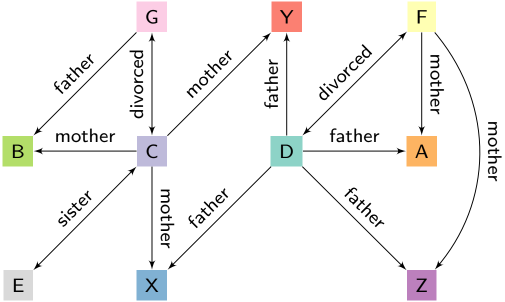
"Applications are submitted for X, aged 1 year, and Y, aged 2 months. The Director of Youth Protection would like X to be entrusted to her aunt, Ms. E, until June 25, 2019. As for Y, that he be entrusted to a foster family for the next nine months. The father has two other children, Z and A, from his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



Redaction in practice

“Applications are submitted for X, aged 1 year, and Y, aged 2 months . The Director of Youth Protection 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 Ms. F. The mother has another child, B, from her union with Mr. G”

☞ background knowledge

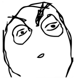


Redaction in practice

Cour de cassation, civile, Chambre civile 1, 10 avril 2013, 12-14.525

“the American company Coca Cola Company markets drinks under the French trade mark "Coca Cola light sango", of which it is the proprietor;”

“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”



Redaction in practice

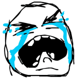
Cour de cassation, civile, Chambre civile 1, 10 avril 2013, 12-14.525

“the American company Coca Cola Company markets drinks under the French trade mark "Coca Cola light sango", of which it is the proprietor;”

“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”

“in this case Abdel X maintains that the patronymic name X enjoys an exceptional reputation since Sango is the language of the Ubangian group of the Republic. Central African, spoken by two million people”

☞ semantics



In the meantime…

- Privacy scandals led to major breaktroughts for publishing structured data privatly
- Differential privacy[6] $\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(y) \in \mathcal{S}]$
  - “An observer cannot tell whether the information from a particular individual were used in the calculation”
  - Combination rules and post-processing
- Differential privacy NLP models[7]

✗ Human readable

✓ Analytics

✓ Privacy

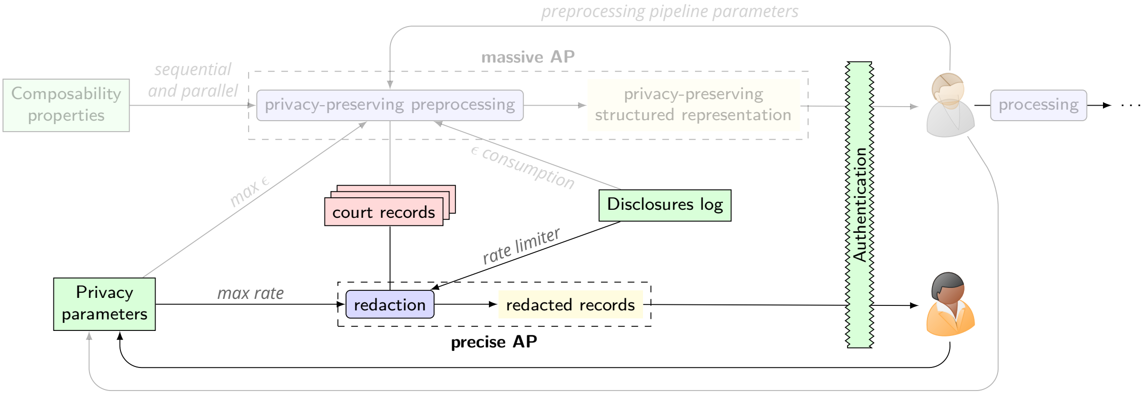
[6] Cynthia Dwork. “Differential privacy”. In: *ICALP*. 2006.

[7] Benjamin Weggenmann and F. Kerschbaum. “SynTF: Synthetic and Differentially Private Term Frequency Vectors for Privacy-Preserving Text Mining”. In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (2018); Natasha Fernandes, M. Dras, and A. McIver. “Generalised Differential Privacy for Text Document Processing”. In: *ArXiv* abs/1811.10256 (2019).

Towards a multimodal publication scheme

☞ Reconcile privacy with transparency by distinguishing two needs:

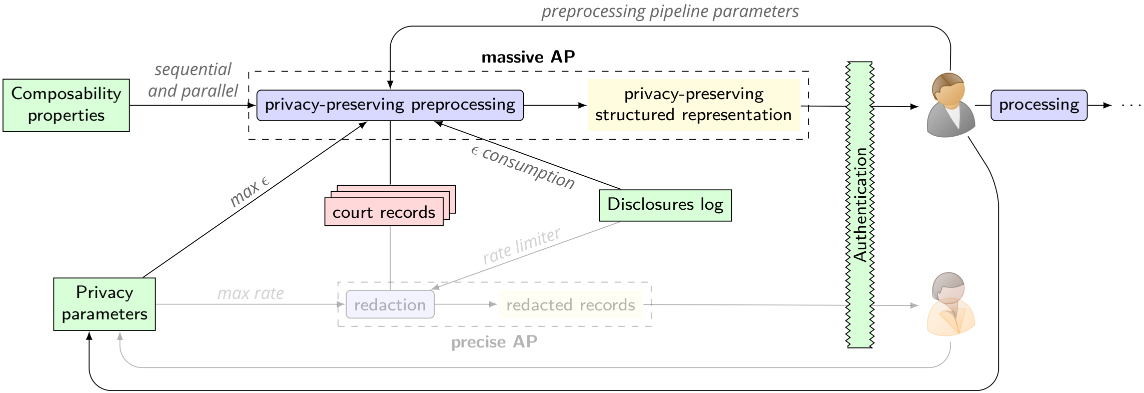
- “Precise” access
- “Massive” access



Towards a multimodal publication scheme

☞ Reconcile privacy with transparency by distinguishing two needs:

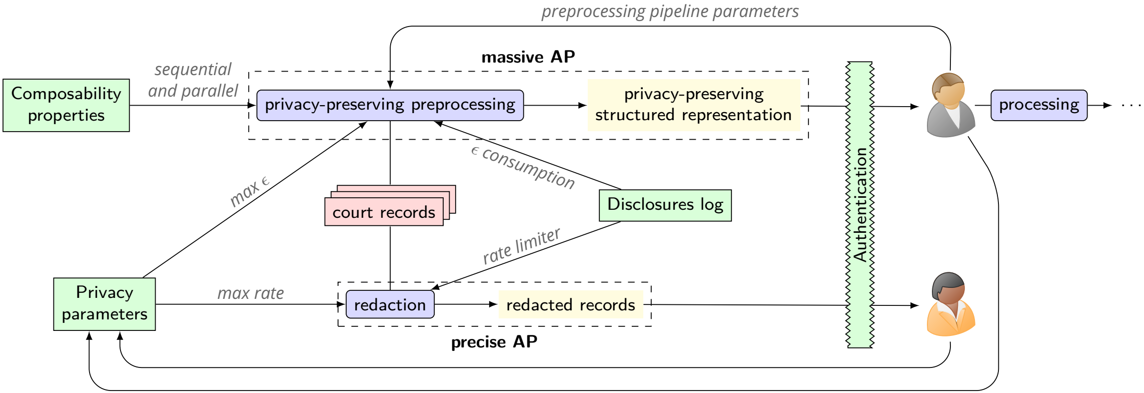
- “Precise” access
- “Massive” access



Towards a multimodal publication scheme

☞ Reconcile privacy with transparency by distinguishing two needs:

- “Precise” access
- “Massive” access



Conclusion

✍ Rule-based redaction is limited

✍ Text anonymization is hard

☞ Discarding the one-size-fits-all approach allows for transparency and privacy

➢ How to reason with composability properties from different models?

➢ Genericity of disclosure logs and real-time access

➢ Express and enforce authentication policies