



Intervention de

Monsieur Jean-Paul Jean

Président de chambre à la Cour de cassation, directeur du Service de documentation, des études et du rapport

« Penser les finalités de la nécessaire ouverture des bases de données de jurisprudence »

Colloque

« La jurisprudence dans le mouvement de l'open data »

Vendredi 14 octobre 2016

Je remercie Mr Petitcollot, président de l'ADIJ (Association pour le développement de l'informatique juridique) d'avoir rappelé le rôle pionnier de la Cour de cassation et du SDER, et je tiens à souligner notamment les initiatives prises il y a plus de 15 ans par l'un de mes prédécesseurs Emmanuel de Givry. La délibération de la CNIL du 29 novembre 2001¹ constitue d'ailleurs toujours la référence concernant les bases de données jurisprudentielles gérées par la Cour de cassation, JuriNet et JuriCA.

Vous l'avez souligné, M. le Premier président, ce colloque marque la volonté de la Cour de cassation de s'inscrire pleinement dans ce mouvement d'ouverture des données publiques. Sous l'égide du Premier président Guy Canivet, et en phase avec les avocats aux Conseils, la Cour avait été aussi pionnière au début des années 2000 dans l'utilisation des nouvelles technologies pour son travail interne, et cela en fait aujourd'hui une des cours suprêmes les plus modernes en Europe dans l'usage des procédures dématérialisées.

Mais ces questions sont posées en des termes totalement nouveaux aujourd'hui dans la perspective de l'open data.

Le mouvement dans lequel vous avez inscrit notre Cour a pour but un accès facilité aux décisions de justice, expression qui revêt un double sens. L'accessibilité, c'est d'abord une évolution de la motivation afin de la rendre plus compréhensible par le citoyen, travail délicat mais qui est engagé en interne et qui produit déjà des effets mesurables.

¹ Délibération CNIL n° 01-057 du 29 novembre 2001 portant recommandation sur la diffusion de données personnelles sur internet par les banques de données de jurisprudence

L'accessibilité c'est aussi une diffusion d'une toute autre ampleur que celle réservée aux cercles d'initiés. L'objectif final fixé est la mise à disposition de toute la jurisprudence à titre gratuit au public, sous réserve de la protection de la vie privée des personnes. La « balance des intérêts » entre ces deux principes, qui constitue la ligne de la jurisprudence des deux cours européennes², va guider tous nos travaux aujourd'hui.

Les décisions de justice, entre données publiques et données personnelles

La question centrale qui se pose est celle des données personnelles contenues dans les décisions de justice, spécificité qui appelle une protection toute particulière, et qui la distingue d'autres documents de justice ou relatifs à la justice n'intégrant pas de données identifiantes.

Pour illustrer cette différence, on peut prendre comme exemple le plus récent la base de données interactive de la CEPEJ que nous venons de rendre accessible au public sur Internet, avec les chiffres, commentaires et analyses détaillés concernant les systèmes judiciaires des 47 Etats du Conseil de l'Europe - soit 850 millions d'habitants -. Nous avons mis en ligne toutes ces données sur les trois derniers exercices 2012, 2014, 2016, ce qui en fait la plus grande base de données du monde relative aux systèmes judiciaires³.

Données personnelles, protection de la vie privée, droit à l'oubli⁴ : c'est donc bien cette problématique essentielle de la protection des données sensibles qui, comme en matière de santé, conditionne tout le reste et qui va être étudiée dans la première partie de ce colloque. La distinction entre les données publiques intégrant des données personnelles et celles qui n'en intègrent pas nécessite une protection renforcée pour les décisions judiciaires qui contiennent les informations les plus intimes. Et la protection spécifique de la vie privée via l'anonymisation, précisée dans la loi Lemaire par le 1^{er} § de l'article L.111-13 du code de l'organisation judiciaire, a été opportunément renforcée, à l'initiative du Sénat, par la volonté de prévenir autant que possible le risque de « ré-identification », dans le §2 du même texte.

Le directeur du SDER est directement responsable de l'effectivité de cette protection pour les justiciables et personnes citées dans les décisions de justice. Mais des quelques cas annuels que nous connaissons actuellement, nous allons passer à une toute autre échelle.

² CEDH, 25 février 1997, Z c/ Finlande, n° 22009/93 ; CJUE, 13 mai 2014, Google Spain SL et Google Inc. c/ Agencia Española de Protección de Datos (AEPD) et Mario Costeja González, aff. C-131/12 ; CJUE, 8 avril 2014, Digital Rights Ireland Ltd c/ Irlande et autres, aff. C-293/12 et C-594/12

³ <http://www.coe.int/t/dghl/cooperation/cepej/evaluation/2016/STAT/default.asp>. NDR : J-P Jean préside le groupe des experts auteur du rapport de la CEPEJ

⁴ http://www.coe.int/t/dghl/cooperation/cepej/newsletter/2015/newsletter_Nov2015_fr.asp

Quelques chiffres permettent de mesurer l'ampleur du changement annoncé.

- La situation actuelle : 500.000 décisions judiciaires anonymisées en accès libre

Le flux de JuriNet. Fin 2017, sur la base du flux actuel nous serons à près de 500.000 arrêts anonymisés accessibles sur JuriNet et Légifrance. En 2015, 10.067 arrêts de cassation et 3.657 arrêts de cours d'appel ont été mis à disposition du public, avec des écarts très importants selon les cours au regard de leurs activités juridictionnelles respectives.

- L'objectif à un an : 1,5 million de décisions judiciaires en plus

Le stock de JuriCa. Si l'internalisation à la Cour de cassation du processus d'anonymisation, avec l'appui d'un logiciel dédié, nous permet d'aboutir à une marge d'erreur ou de doute faible pour un contrôle ciblé avec intervention humaine résiduelle pour les levées de doute, nous pourrions effectivement mettre à disposition du public d'ici fin 2017, conformément à l'objectif fixé par le Premier président Louvel, 10 années d'arrêts anonymisés des cours d'appel stockées dans JuriCa, soit 1,5 million de décisions non pénales rendues en audience publique (environ 150.000 sont rendues chaque année).

Pour la Cour de cassation, l'ampleur du défi à relever est conditionné par les performances évaluées du logiciel actuellement en phase de test et le renforcement des moyens avec l'appui de la direction des services judiciaires pour répondre à cet objectif. Mais le travail d'anonymisation dépend aussi du niveau d'exigences posé, qui est élevé pour la Cour de cassation, d'où l'importance des normes fixées par la CNIL.

Nouveauté aussi, mais essentielle, la prévention du risque de ré-identification, d'où l'enjeu du décret d'application de l'article L.111-13 du code de l'organisation judiciaire, pour savoir qui identifie ce risque et selon quelles modalités.

- L'objectif final : 1,5 million de décisions judiciaires de plus chaque année

La réussite de cette démarche progressive est déterminante pour arriver à ce que vous souhaitez à terme, M. le Premier président Louvel, pour toutes les juridictions, à savoir dans une seconde phase, la mise à disposition du public sous forme anonymisée de toutes les autres décisions de justice rendues publiquement, et ne bénéficiant pas d'une protection particulière. Cela concerne les arrêts de juridictions pénales des cours d'appel et les décisions civiles et pénales des juridictions de première instance et implique donc des choix du ministère de la Justice via les grands programmes informatiques Portalis et Cassiopée, les trames des jugements, la formation des personnels de greffe etc.... Le travail que nous effectuons en commun avec le ministère, plus particulièrement la direction des services judiciaires, fait aussi de la Cour de cassation le garant de la protection des droits individuels des personnes citées dans les procédures juridictionnelles.

Cette perspective constitue un objectif mesurable, que permettra l'évolution technologique selon les choix et les priorités d'investissement du ministère de la justice. Dans quelques années cela représenterait plus d'1,5 million de décisions anonymisables chaque année :

- 105.000 arrêts pénaux de cour d'appel (45.000 arrêts correctionnels, 40.000 arrêts de la chambre de l'instruction, 20.000 arrêts des chambres de l'application des peines) ;
- Environ 600.000 jugements des tribunaux correctionnels ;
- Environ 830.000 jugements (hors référés, soit 380.000 décisions du juge aux affaires familiales, 90.000 décisions du juge de l'exécution, 100.000 décision du juge des libertés et de la détention).

A terme, le changement d'échelle est donc considérable ; les moteurs de recherche vont tourner et cette matière grise précieuse va produire une course entre explorateurs, avec une concurrence farouche pour découvrir des pépites avec l'appui des algorithmes.

Pour avancer rapidement, avec toutes les garanties nécessaires, nous sommes ouverts à aider l'élaboration d'un logiciel libre de type « *machine learning* ». C'est pourquoi par exemple, nous avons aussi passé une convention de recherche avec la start-up doctrine.fr, car soutenir l'innovation et la concurrence sont ici les meilleurs moteurs pour progresser.

Une fois cette perspective fixée, je souhaite, en vue des débats qui vont suivre, poser quelques questions qui me paraissent importantes.

L'open data, entre service public et marché

Toutes ces décisions présentent-elle un intérêt pour la diffusion de la jurisprudence ? Est-il plus simple de tout traiter de façon quasi-automatisée, ce qui fera le bonheur des start-ups qui feront tourner les logiciels et les algorithmes de plus en plus sophistiqués, sur des niches d'information les plus diverses qui intéresseront des clients pour un service payant correspondant à des besoins spécifiques, qui peuvent ne pas être uniquement juridiques ?

Ou bien, est-il souhaitable et peut-on réellement cibler ce qui constitue la finalité première, la diffusion de la jurisprudence pour aider à la régulation par le droit, et donc sélectionner les décisions ? Le comité des ministres du Conseil de l'Europe, dans une recommandation de 1995 préconisait d'ailleurs une sélection des décisions afin « d'éviter l'accumulation d'informations inutiles⁵ ».

Le fait qu'une décision judiciaire soit rendue en audience publique est une chose, qu'elle soit remise aux parties en est une autre, qu'elle soit diffusée publiquement et notamment sur Internet, en est une troisième ; qu'elle soit retraitée avec des millions d'autres via les algorithmes des moteurs de recherche nous fait entrer dans une nouvelle dimension dont

⁵ Recommandation n° R (95) 11 du comité des ministres aux Etats membres relative à la sélection, au traitement, à la présentation et à l'archivage des décisions judiciaires dans les systèmes de documentation juridique automatisés

nous ne pouvons pas mesurer aujourd'hui l'ampleur des conséquences. Les services proposés se démultiplient aujourd'hui dans le marché de l'information juridique et judiciaire, bien au-delà des seuls domaines de la banque ou de l'assurance, et plus généralement chez les acteurs intéressés à l'évaluation des risques par les calculs de probabilité et l'analyse actuarielle.

Dans ces nouveaux services, les **services publics** peuvent conduire des projets. Ainsi aux Pays-Bas le projet récompensé par la CEPEJ "Résolution de conflits en ligne pour les conflits relationnels" consiste en une plate-forme en ligne intégrant un dispositif d'accès à l'aide judiciaire, l'orientation vers une procédure négociée et un lien avec le conseil juridique fourni par un avocat spécialisé⁶.

Mais c'est surtout le **marché** qui se développe sur trois types de services :

- ceux liés à la recherche juridique avec de nouveaux moteurs de recherche de décision de justice, plus rapides et plus simples⁷ ;
- les services proposant des alternatives à un procès, tel que l'arbitrage en ligne, qui existe déjà⁸.
- les services facilitant les démarches liées à un procès, telles que la recherche d'un avocat ou la mise en contact avec un avocat spécialisé⁹, ou l'aide à la constitution d'un dossier et à la saisine d'une juridiction¹⁰.

D'ores-et-déjà, à partir des données disponibles, les **éditeurs juridiques** proposent des outils de prédiction chiffrée des décisions de justice, à travers la compilation de « jurisprudence chiffrée » sur certains thèmes particuliers¹¹ et la recherche de précédents en « contentieux de l'indemnisation »¹².

La révolution de l'open data va développer et faire évoluer ce type de services, en permettant, dans certains domaines, d'analyser non seulement la jurisprudence d'une juridiction, mais aussi celle d'une chambre par rapport à une autre, d'un juge unique ou d'un juge rapporteur en collégialité. Nous avons déjà vu en France des analyses de taux de décisions allant dans un sens ou un autre par juge au sein d'une même juridiction administrative¹³. Même si la méthodologie de cette étude est discutable, le seul fait qu'elle

⁶ Bureau d'Aide judiciaire des Pays-Bas (Raad voor Rechtsbijstand), projet "Résolution de conflits en ligne pour les conflits relationnels" <https://www.youtube.com/user/Rechtwijzer>

⁷ Ainsi le moteur de recherche développé par www.doctrine.fr

⁸ Le site www.ejust.fr propose le règlement par l'arbitrage en ligne des litiges commerciaux

⁹ Le site www.flash-avocat.fr permet de transmettre des documents à un avocat pour solliciter un premier avis

¹⁰ Le site <https://www.demanderjustice.com/> propose pour certains domaines sélectionnés l'envoi d'un dossier au tribunal après une première phase de tentative de résolution amiable du litige

¹¹ Service « Jurisprudence chiffrée » de Dalloz (<http://jurisprudencechiffree.dalloz.fr/>) et Francis Lefebvre (<http://boutique.efl.fr/documentation/par-gamme/jurisprudence-chiffree.html>)

¹² Service « Contentieux de l'indemnisation » de LexisNexis (http://www.lexisnexis.fr/solutions/infocherche/Contentieux_Indemnisation/)

¹³ Michaël Benesty, *L'impartialité de certains juges mise à mal par l'intelligence artificielle*, <http://www.village-justice.com/articles/impartialite-certains-juges-mise,21760.html>, 24 mars 2016

existe constitue un indice, « un signal faible » à ne pas ignorer, nous apprend la démarche prospective.

Là encore, il ne s'agit pas d'un changement de pratique, mais d'un changement de dimension dans le rapport au juge ou à l'avocat, que va apporter l'open data.

Au Canada, il existe une devise pour apprécier la compétence d'un avocat : “*A good lawyer knows the law, an excellent lawyer knows the judge*”.

Cela ne veut pas dire que l'avocat connaît personnellement le juge, mais qu'il connaît sa jurisprudence, sa façon d'appréhender un contentieux et qu'il va donc chercher à le convaincre par des arguments pertinents auxquels il sait que ce juge est sensible¹⁴.

Mais cela veut aussi dire que, pour gagner son procès, il va plutôt chercher à faire orienter son affaire vers ce juge, ou tenter de le contourner, en recourant au *forum shopping*, en jouant sur les règles de compétence territoriale, sur les attributions de différentes chambres d'un tribunal, voire sur les dates d'audiences et les demandes de renvoi, pour lesquelles aussi il faut parfois connaître le greffier autant que le juge...

Face à ces pratiques de contournement, seul le principe du juge naturel et les systèmes d'attribution des dossiers automatisés permettent de répondre.

Le big data implique-t-il une anonymisation renforcée ?

C'est pour anticiper ces questions qu'il faut **évaluer les avantages et les risques liés à l'identification des professionnels, magistrats et avocats.**

La plus grande prévisibilité des décisions de justice, les rapprochements de jurisprudence, conduisant naturellement à une atténuation des éventuelles divergences, à une autorégulation des jurisprudences sous l'autorité de la Cour de cassation dans son rôle unificateur et normatif, s'intègre parfaitement dans un objectif d'harmonisation de l'application du droit.

Pour autant, la question de la prédictibilité peut être envisagée sous un autre angle : celle de la prédictibilité des résultats des professionnels pris un par un. Si je confie mon affaire à tel ou tel avocat, vais-je gagner ou perdre mon procès ? Vers quel juge dois-je tout faire pour orienter mon affaire afin d'obtenir gain de cause ?

Aux États-Unis, le site internet *Lex Machina* (*Lexisnexis company*) vient d'annoncer le lancement d'un nouveau service : « le comparateur de juges et de tribunaux » et le « comparateur de cabinets d'avocats »¹⁵.

¹⁴ Un avocat peut déjà prendre tout renseignement sur les jurés par les informations directement accessibles sur Internet, en particulier celles accessibles via les réseaux sociaux, soit pour les récuser, soit pour choisir les arguments auxquels certains pourraient être sensibles. Cf. Pascale Robert-Diard, *Choisir ses jurés sur internet*, JCP, Ed. Gén., n° 41, 10 octobre 2016, p 1841

¹⁵ <https://lexmachina.com/>

Le fondateur du site explique : « Maintenant les avocats ont le pouvoir de comparer les tribunaux, les juges, et les cabinets en un clic. L'avocat en demande peut désormais utiliser l'application pour comparer différents tribunaux avant d'enregistrer sa requête, afin de sélectionner celui qui apportera probablement l'issue la plus favorable dans les meilleurs délais ».

Le site *Ravel Law* (collaboration avec l'Université d'Harvard) propose un service similaire d'analyse des juges, qui promet aux internautes de « comprendre comment [leur] juge pense, écrit et décide »¹⁶.

Le site *Court Listener* (en open source) permet quant à lui de faire des recherches sur un juge, pour connaître son parcours, son éventuelle affiliation politique et les conditions de sa nomination¹⁷.

Cette démarche, parfaitement naturelle aux États-Unis, où de nombreux juges sont élus, peut-elle être importée en France ? N'y-a-t-il pas des risques pour certains juges, notamment ceux ayant à connaître des affaires de terrorisme ou de grande criminalité ?

Le risque de ré-identification

Il convient aussi de mesurer les risques liés à l'identification des parties et témoins au procès. L'article 21 de la loi pour une République numérique intègre ces impératifs en précisant, d'une part, que les décisions de justice sont mises à disposition « dans le respect de la vie privée des personnes concernées » et, d'autre part, que « cette mise à disposition du public est précédée d'une analyse de risque de ré-identification des personnes ».

L'on sait qu'il sera possible d'identifier les personnes concernées par une décision de justice publiée sur internet. En effet, même si certaines mentions sont occultées (par exemple, le nom des personnes physiques), le numéro de RG (registre du greffe) qui figure sur chaque décision permettra à toute personne de demander au greffe la communication de la décision non anonymisée.

Par ailleurs, de nombreuses décisions, et plus encore celles de décisions des juges du fond, contiennent tant de détails liés aux faits de l'espèce que des personnes connaissant les protagonistes d'un procès pourront les reconnaître à la lecture d'une décision, même si les éléments d'identification directe sont occultés.

L'enjeu ne saurait donc être d'empêcher toute ré-identification, ce qui est impossible (à moins de rendre les décisions illisibles et inexploitables), mais d'évaluer ce risque, de le limiter, donc de définir un niveau de risque acceptable, doublé d'un mécanisme de correction, au regard des bénéfices attendus par l'open data.

¹⁶ <http://ravellaw.com/products/>

¹⁷ <https://free.law/tag/courtlistener.html>

Est-il acceptable qu'un employeur puisse vérifier si un candidat à l'embauche a connu des litiges civils ou des condamnations pénales, avec en outre tous les détails : a-t-il été condamné pour conduite en état alcoolique ? A-t-il déjà attrait un employeur devant le conseil des prud'hommes ? Est-il mauvais payeur ? Connaît-il des difficultés familiales susceptibles d'affecter son engagement professionnel ?

Répondre à ces questions, c'est ce que propose aux Etats-Unis le site internet *checkr.com*, qui compte d'ailleurs plusieurs Français parmi ses dirigeants. Ce site internet offre de simplifier la recherche d'antécédents d'un salarié, ce qui constitue parfois, aux Etats-Unis, une obligation pour l'employeur, susceptible d'engager sa responsabilité si elle n'est pas effectuée. Ce service aboutit à la rédaction d'un rapport après consultation, automatisée ou non, des registres de sécurité sociale, des fichiers des agresseurs sexuels, et des décisions de justice pénales et civiles disponibles¹⁸.

Au regard des règles européennes, mais aussi tout simplement du système de valeurs qui doit être le nôtre, ce type de service n'est évidemment pas acceptable.

Penser les finalités de l'ouverture des bases de données de jurisprudence

Au regard du droit des données personnelles, un traitement de données doit être attaché à une **finalité précise**. Le règlement européen sur la protection des données personnelles prévoit ainsi dans son article 5 un principe de « minimisation des données » : « les données à caractère personnel doivent être [...] adéquates, pertinentes et limitées à ce qui est nécessaire pour les finalités pour lesquelles elles sont traitées »¹⁹.

Si l'on s'interroge sur la finalité des bases de données juridiques, il s'agit certainement de permettre une meilleure connaissance et une meilleure transparence des règles de droit et de leur application à des circonstances de fait. La simple connaissance des formations de jugement permet de comparer et de rapprocher les jurisprudences. La collecte massive d'informations sur la personne des juges, des greffiers ou des avocats, permettant de générer des statistiques sur l'ensemble de leur activité professionnelle et de leurs « performances », voire de leur système de valeurs (nombre d'affaires par an, moyenne de temps de traitement, résultats, motivation, etc...) n'intègre pas cette finalité.

En conclusion, le message est clair selon lequel la volonté d'ouverture de la Cour de cassation est totale pour que la diffusion de la jurisprudence soit la plus large possible.

En tant que **producteur de données**, il convient cependant de conserver ce qui fait l'essence de la Cour de cassation, à savoir la valeur ajoutée qui résulte non seulement de l'indispensable anonymisation mais aussi de l'enrichissement de décisions pour éviter une

¹⁸ <https://checkr.com/privacy-policy/>

¹⁹ Règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données)

masse informe qui ne permettrait plus de distinguer l'important de l'insignifiant, qui noierait le juriste et le citoyen dans une marée de décisions où un moteur de recherche établirait ses propres hiérarchies, à finalités parfois discutables. D'où l'importance du travail à la source, sur un original certifié et dont le respect de l'intégrité peut toujours être aisément vérifié, de la hiérarchie des arrêts publiés par un mode d'identification clair et connu des juristes, avec un enrichissement par des sommaires et des titres, les références doctrinales.

L'open data nous pose un **défi quantitatif**. Nous devons y répondre mais renforcer dans le même temps le **défi qualitatif** de la production à la source.

Jean-Paul Jean