

L'enjeu de l'*open data* pour les éditeurs et les *legaltech* : une redéfinition du marché de l'information juridique

Sumi Saint Auguste, Responsable du Lab Prospective, groupe Lefebvre Sarrut - Présidente de l'association Open Law* Le droit ouvert

Bien sûr, il y a le *Moniteur universel* de Charles-Joseph Panckoucke qui, publiant les débats parlementaires dès 1789, préfigure le *Journal officiel*. Bien sûr, il y a depuis le XIX^e siècle les multiples libraires-éditeurs du V^e arrondissement qui, adossés à l'Université, ont fourni en manuels des générations d'étudiants en droit. Mais si l'on devait dérouler le récit de cette aventure intellectuelle et industrielle qu'est l'édition juridique française, on serait tenté de convoquer une figure, emblématique, qui est celle de Désiré Dalloz et de son Recueil éponyme, ou encore son pendant de la justice administrative : le Lebon. En effet, les commentaires d'arrêts et l'analyse de la jurisprudence fondent historiquement le cœur de métier et la raison d'être des éditeurs juridiques en France. L'éditeur est en effet l'acteur et l'opérateur-clé de la jurisprudence qui s'écrit et se construit, et son rôle auprès des auteurs de doctrine est fondamental.

Avec le basculement en flux gratuit des licences Légifrance en 2014, le principe d'*open data* public fait son entrée dans le champ de l'information légale et marque le début d'une discrète et lente révolution appelée à modifier de façon irréversible l'économie de l'information juridique. Comme dans tous les autres secteurs que l'ouverture et la mise à disposition massive de données publiques redessinent radicalement, cette irruption de l'*open data* suscite l'émergence de nouveaux acteurs intéressés par les perspectives de réutilisation des données publiques. Elle éclaire aussi, sous un jour désormais différent, le rôle des acteurs historiquement présents que sont les éditeurs juridiques, bousculant autant les modèles économiques que leur proposition de valeur.

I - Des enjeux de qualité de l'information juridique

Depuis que la loi pour une République numérique en a posé le principe en octobre 2016, l'*open data* des décisions de justice a fait apparaître une équation nouvelle entre des principes *a priori* inconciliables. L'*open data* consiste en effet en une mise à disposition indifférenciée, c'est-à-dire que la donnée est rendue disponible pour le plus grand nombre, sans préjuger de l'usage qui en sera fait. Il importe en conséquence que tout élément visant à identifier des personnes physiques soit impérativement occulté afin de garantir une protection absolue des données personnelles avant toute diffusion.

A - La question de l'intégrité des données

Ces éléments identifiants nourrissent nécessairement le récit des décisions de justice qui mêlent des éléments de fond autant que de droit. La décision ainsi tronquée perd potentiellement toute lisibilité et, les empêchant d'appréhender aisément l'information, se dérobe au rédacteur ou à l'auteur. Cette question d'une donnée non dégradée a très tôt alimenté un vif débat et un point d'alerte pour les éditeurs autant que pour les avocats qui ont fait valoir auprès du législateur, puis de la Chancellerie, la nécessité de maintenir en parallèle de l'*open data* un « flux de données intègres », réservé à un usage professionnel dûment identifié. La lisibilité des décisions où les noms des parties ont été remplacés par des initiales continue d'être discutée et fait l'objet de démarches des acteurs du marché.

B - Le défi industriel d'une mise à disposition massive

Le rôle échoit aux cours suprêmes d'organiser désormais cet *open data* des décisions de justice, là où précédemment il était confié, pour l'ordre judiciaire, à la Direction de l'information légale et administrative (DILA). Outre l'impératif absolu de l'anonymisation, la qualité de la donnée intègre un impératif de fraîcheur et de rapidité du délai de mise à disposition du flux en *open data*, qui représentent un défi conséquent en termes de chantier informatique. Celui-ci est étroitement dépendant de la modernisation des systèmes d'information organisés par la Chancellerie, assurant la collecte, la transformation et le transport jusqu'à l'opérateur de l'*open data* judiciaire qu'est aujourd'hui la Cour de cassation. Il faut avoir à l'esprit que ce projet informatique d'envergure entraîne une modernisation accélérée des systèmes d'information du pouvoir judiciaire, tant dans les chaînes de collecte que dans la publication sur internet. Ainsi, l'*open data* s'accompagne, pour le site de la Cour de cassation, de la migration vers un nouveau moteur de recherche plus performant.

Parce que l'*open data* implique une modernisation radicale des moyens et des circuits de production de cette donnée publique, on observe que le bénéfice dépasse l'enjeu strict de cette stratégie d'ouverture et entraîne une optimisation des flux de production, possiblement aussi de l'organisation et des rôles. L'*open data* implique enfin le recours à des technologies et outils nouveaux, avec par exemple le recours à des algorithmes de *machine learning* pour anonymiser les décisions et, de façon corollaire, de nouvelles interfaces d'annotation pour contrôler et fiabiliser la qualité des décisions de justice ainsi traitées par le moteur d'anonymisation : la technologie joue ici encore pleinement son rôle d'accélérateur d'innovation.

II - Des enjeux de marché

A - Des rapports public/ privé redéfinis

L'*open data* a commencé par redéfinir les rapports entre ces acteurs publics et privés.

Ce qui a encore cours dans d'autres pays européens où la diffusion est déléguée au secteur privé, a commencé à être bouleversé avec la loi pour une République numérique d'octobre 2016.

Aux côtés de la DILA, l'association Open Law* Le droit ouvert a organisé début 2017 un programme d'innovation ouverte composé de tables rondes et d'ateliers techniques, qui a fédéré les différents acteurs du secteur. Ce programme se voulait un espace de dialogue entre acteurs du secteur marchand et institutions, exempt de la dynamique de lobbying ou d'influence attendus ailleurs. La réunion de la DILA, des cours suprêmes, de la CNIL, de l'Institut national de la propriété industrielle (INPI) et de nombreux éditeurs a été rendue possible car les objectifs communs visaient la définition, voire la construction, d'outils et de ressources concrets utiles à tous, comme notamment la cartographie du circuit des données judiciaires et administratives ou l'enrichissement des vocabulaires nécessaires pour la norme ECLI.

On soulignera aussi que la mise en oeuvre de cet *open data* judiciaire et administratif incombant désormais aux cours suprêmes, le dialogue jadis centré autour de la DILA et des services du Premier ministre autour des aspects d'ingénierie et de production, se déporte désormais vers d'autres interlocuteurs que sont le service de documentation de la Cour de cassation, qui pilote le chantier de l'*open data* judiciaire, ou encore les services informatiques du Conseil d'État.

B - Une reconfiguration du marché et de la concurrence

Le marché de l'édition juridique est avant tout un marché qui s'est construit historiquement autour de la donnée

publique. Avec la refonte complète du site Légifrance en 2020, et le nouveau site de la Cour de cassation annoncé pour septembre 2021, on peut dire que ce marché de l'information juridique et judiciaire trouve son premier compétiteur en l'acteur public, avec deux avantages concurrentiels que sont la gratuité, et bientôt la primauté de la publication sur un même périmètre de données, là où précédemment le décalage était transparent pour le lecteur final, puisque les arrêts de cour d'appel n'étaient dans le champ public qu'à hauteur de quelque 4 % du stock réel.

C - Un déplacement de la chaîne de valeur des éditeurs

Les questions de volume, d'exclusivité et d'intégrité de la donnée ne seront, à terme, plus un avantage concurrentiel dont se prévaloir, ce qui vient bousculer sévèrement le modèle économique de certains acteurs dont la valeur et le positionnement sur le marché se sont construits à partir de ces avantages concurrentiels.

Les éditeurs juridiques se trouvent ainsi à un tournant de leur mission où ils sont incités à réaffirmer, d'une part leur rôle de facilitateur et d'orchestrateur de la doctrine juridique, et pour cela à se recentrer sur leur valeur ajoutée, qui réside dans le mieux-disant en matière d'ergonomie et de fiabilité de la recherche juridique ; d'autre part dans l'ingénierie éditoriale qui tendrait à mobiliser différemment ces données de la jurisprudence pour proposer des éclairages différents, outillés par exemple par des démarches statistiques, et ainsi produire une information juridique innovante.

1 - Les legaltech changent la donne

Les *legaltech* sont des *start-up* qui ciblent le marché du droit : information ou prestations de services à teneur juridique, hors conseil, et s'emploient majoritairement à « plateformiser », autrement dit à appliquer les technologies et fonctionnalités des plateformes à des services juridiques pour des clients, entreprises ou particuliers. Il existe une poignée d'acteurs dans ce nouvel écosystème qui, tels Predictice ou Doctrine, se sont positionnés sur le marché de l'information juridique et de l'accès à la donnée jurisprudentielle, se plaçant en concurrence frontale avec les éditeurs juridiques. L'apparition de ces nouveaux entrants a pour effet mécanique de redistribuer les forces en présence sur le marché.

En 2017, l'apparition du moteur de recherche doctrine.fr a fortement mobilisé les éditeurs qui faisaient face à un nouvel acteur offensif et frondeur. Malgré le contentieux qui oppose cette *start-up* au Syndicat national de l'édition, les relations de défiance ont cédé la place à une cohabitation relativement clémente, l'enjeu pour les éditeurs historiques s'appréciant aussi en termes d'image et de positionnement sur l'innovation notamment technologique. La couverture médiatique et les éléments de langage des protagonistes ont contribué à polariser le débat, dépeignant les uns en acteurs oligopolistiques défenseurs d'un *statu quo* dont ils seraient les bénéficiaires exclusifs, et les autres en chantres du progrès, de l'ouverture et de l'innovation pour le bénéfice de ses clients professionnels du droit. D'une certaine façon, l'épisode Doctrine n'est qu'un jalon supplémentaire dans l'histoire du jeune marché de la *legaltech*, déjà émaillé de quelques polémiques qui ont vu s'opposer de jeunes entreprises innovantes et les instances ordinales de professions réglementées, par exemple autour des services d'annuaire ou de mise en relation avec des avocats, ou ici un nouvel éditeur qui ne dit pas son nom, et ses prédécesseurs. Mais du fait de leur poids économique et culturel dans l'économie de l'information juridique, les confrontations qui ont pu opposer les éditeurs à cette *start-up* ont occupé une place disproportionnée dans la genèse de ce marché *legaltech* émergent.

Au-delà des développements polémiques, cet épisode récent dans l'histoire de l'édition juridique a objectivement contribué à précipiter des choix de développement, ou des migrations technologiques, à entraîner les éditeurs présents dans un jeu de compétition radicalement renouvelé, le dernier entrant ayant pour ainsi dire servi d'aiguillon.

C'est ainsi que l'*open data* et les opportunités qu'il a entraînées, ont activé la mue paradoxale de tout un secteur qui, tout en étant déjà fortement numérisé, a amorcé une véritable transformation digitale.

2 - Recherche documentaire et expérience utilisateur

L'expertise des éditeurs intègre nativement une ingénierie documentaire et un art (au sens de technique) de l'indexation et de la classification des décisions. Avec les premières mises en base de données électroniques, avec les technologies des moteurs de recherche et les ontologies attachées, cette économie de la donnée structurée caractérise très tôt l'édition juridique autant que l'édition scientifique, technique et médicale en matière d'informatique éditoriale et de numérisation du document. À l'heure où de nouveaux acteurs, forts d'une culture numérique dès l'origine, pour qui les bonnes pratiques de recherche et d'ergonomie se situent plutôt du côté des moteurs de recherche grand public, de l'e-commerce et, de façon générale, des plateformes de services en ligne, cet héritage industriel et cette expertise de longue date peuvent se révéler un frein pour des acteurs spécialistes de la recherche experte dont les fonctionnalités ont été conçues pour les initiés. L'ouverture du marché qu'a entraînée l'annonce de l'*open data* achève de rebattre les cartes en matière d'ergonomie, de fluidité dans le parcours de recherche jusqu'aux résultats, la référence étalon étant soudain celle de la recherche naïve opérée dans Google.

3 - De nouveaux modes exploratoires de la jurisprudence

L'annonce de l'*open data* a aussi ouvert le champ des approches probabilistes qui misent sur la disponibilité, du fait de l'*open data*, d'un grand volume de données rendues aisément exploitables par une machine : les bonnes pratiques d'*open data* préconisent en effet de mettre la donnée à disposition dans un format non propriétaire et lisible par une machine (*machine readable*). C'est ainsi que des démarches visant à dégager des tendances ou des motifs récurrents dans les décisions rendues ont été facilitées par l'*open data*. Ici encore, ce n'est pas tant l'effectivité de cet *open data*, qui reste encore très théorique pour les juridictions du premier degré, que la perspective d'une ouverture prochaine qui a suscité un intérêt accru pour l'analyse quantitative des décisions et de leurs motivations. Ces approches préexistaient à l'*open data*, et étaient déjà développées sous la forme de produits ou de fonctionnalités chez certains éditeurs qui ne sont pas sans évoquer l'approche par le niveau de preuve (*evidence-based*) à l'oeuvre en médecine clinique. Il s'agit de recourir à un étayage statistique censément objectif pour poser ici un diagnostic ou un protocole thérapeutique, motiver là une décision ou alimenter des conclusions.

Il y a dans l'approche par le nombre, ou pilotée par la donnée (*data-driven*) la promesse d'objectiver des choix, et de rationaliser en cela le processus judiciaire.

Les acteurs publics autant que les éditeurs, et des *legaltech* comme Case Law Analytics ou Jurispredis, se sont saisis de cette opportunité et doivent nécessairement parer à une forme de défiance quant au risque de barémisation de la justice pour certains, quand il s'agit par exemple, à partir d'une analyse statistique d'une masse de décisions sur un contentieux donné, d'en modéliser les observations et d'en dégager un référentiel : c'est le reproche qui a été fait à l'outil Datajust sur l'évaluation des préjudices corporels en mars 2020. La frilosité observée face à ces outils s'attache majoritairement au respect de la liberté de juger des magistrats, et particulièrement leur créativité, quand justement on établit que les modèles statistiques figent par définition des décisions passées. Il est intéressant de constater que le législateur, autant que les acteurs présents, ont déjà activement déroulé les hypothèses d'exploitation de cette donnée à des fins prédictives, et apprécié les risques de dérives, notamment à la lumière de l'abondante littérature sur les biais statistiques, alors même que le stock d'arrêts et de jugements n'est pas encore largement ouvert.

III - Au-delà de l'*open data* des décisions de justice

De même que l'on a pu souligner son effet de levier et d'accélérateur de la modernisation des outils et chaînes de

production de la donnée de jurisprudence, l'*open data* a d'une certaine façon pour mérite de questionner activement les usages et l'économie de ladite donnée, et partant d'interroger le rôle des éditeurs et des acteurs de l'information juridique. Cette doctrine de l'*open data* trouvait ses fondements dans le rapport Trojette⁽¹⁾ sur l'ouverture des données publiques en 2013 ; elle se trouve à nouveau démontrée et promue dans le rapport Bothorel de décembre 2020⁽²⁾ sur la politique publique de la donnée. Dans ses recommandations, le rapport de la mission Bothorel alerte notamment sur la garantie d'ouverture des données publiques qui seraient mises en oeuvre par des acteurs privés dont des professionnels de justice. On peut y lire une préconisation plutôt volontariste à propos des décisions et tout autant des données enregistrées par les tribunaux de commerce. Ces données surabondantes du contentieux des entreprises deviennent, dans le contexte de la récession économique, soudain attractives pour des éditeurs et *start-up legaltech* qui pourraient utilement développer des solutions à partir de ces données sur les entreprises. À l'instar par exemple de l'application Signaux faibles développée par Etalab, l'opérateur de la donnée publique au sein de la Direction numérique de l'État, pour le compte du ministère des Finances, et qui vise à détecter précocement les entreprises en difficulté, on peut imaginer la variété des scénarios d'utilisation pour le bénéfice de tous les acteurs économiques.

On peut aisément considérer que ce nouvel épisode de l'*open data* public pourrait occasionner une redistribution des acteurs économiques présents sur cette économie de la donnée relatives aux entreprises. Tout comme à propos des données de la jurisprudence, des *start-up* telles que Pappers, Rubypayeur ou Alexlegal ont entrepris de chahuter le modèle économique des opérateurs traditionnellement présents, à commencer par celui du groupement d'intérêt économique Infogreffe, en proposant gratuitement des moteurs de recherche sur les données disponibles en *open data* : base SIRENE, avis du BODACC, etc.

Sur les données de justice comme pour le réservoir des données légales des entreprises, on constate à quel point ces enjeux d'*open data* s'apprécient en termes de transparence et d'accès démocratique au droit, pour tous les professionnels du droit, comme pour les citoyens ; en outre, ils résonnent désormais avec un impératif plus conjoncturel d'efficacité économique renforcée : la reprise économique des années à venir en dépend.

Mots clés :

INFORMATION JURIDIQUE * Edition * Legaltech

(1) Rapport Trojette sur l'ouverture des données publiques, publié le 5 nov. 2013.

(2) Le rapport sur la politique publique de la donnée, des algorithmes et des codes sources, a été remis par Éric Bothorel au Premier ministre le 23 déc. 2020.