

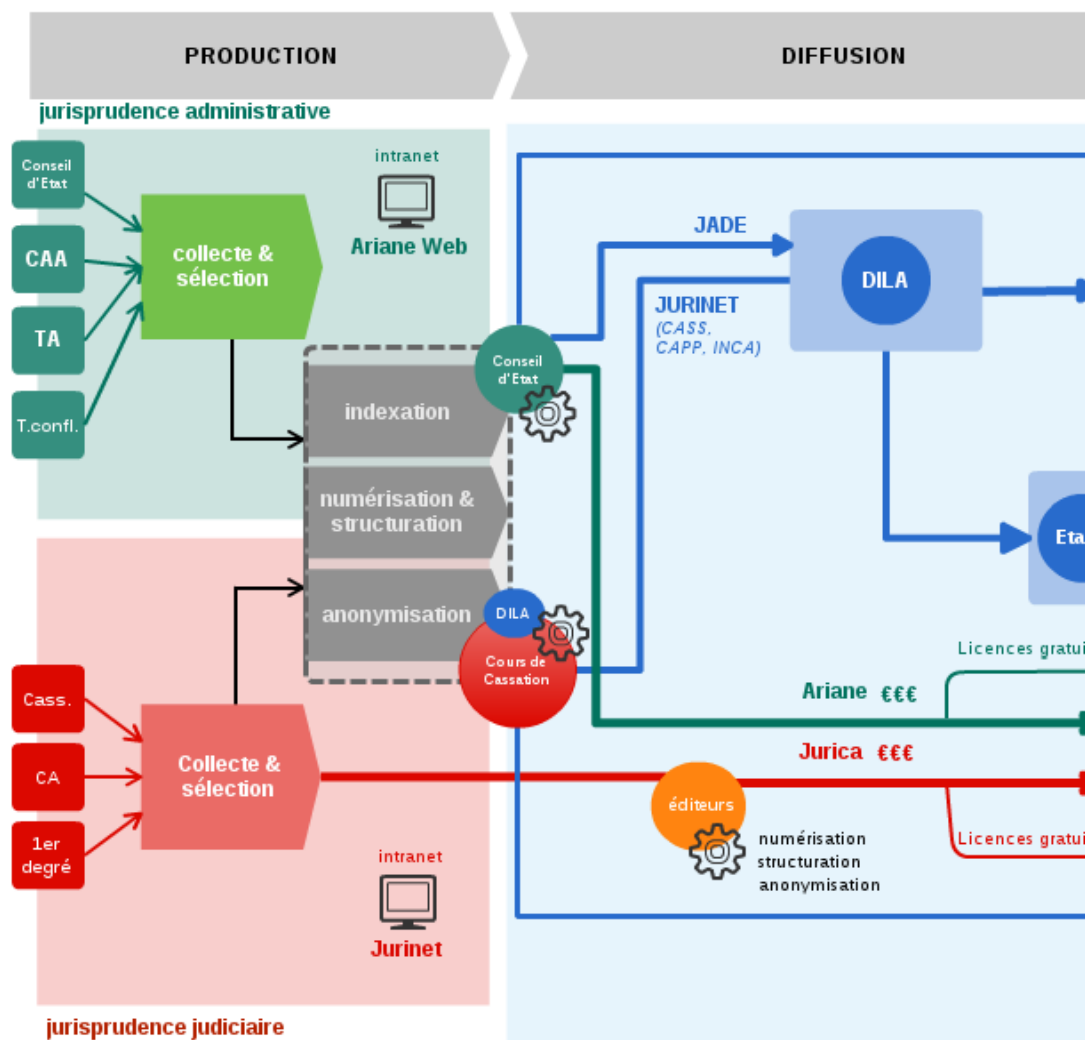
Livre Blanc sur l'open data jurisprudentiel

Synthèse des travaux du programme

Open «Case» Law - Janvier 2017

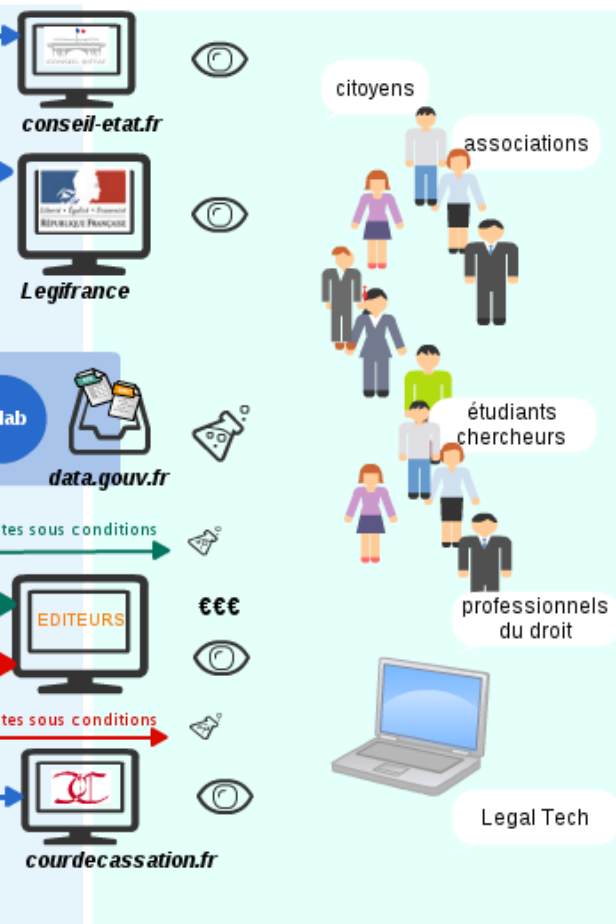
**OPEN
LAW***

* Le droit ouvert



UTILISATION

*De données
jurisprudentielles*



Format PDF
non anonymisé



Format XML
anonymisable +
Format word anonymisé
Indexation et analyses



format XML
anonymisé +
Métadonnées



Solution technique
de numérisation
& d'anonymisation



open access



open data



Le 15 septembre 2016, la DILA et l'association Open Law ont lancé un programme court sur l'ouverture des données de jurisprudence en partenariat avec Etalab, le ministère de la Justice, la Cour de cassation, la Cour des comptes, le Conseil d'État, l'Ordre des Avocats de Paris, la CNIL et l'INPI¹.

L'adoption des articles 20 et 21 de la Loi pour une République numérique du 7 octobre 2016² prévoyant l'ouverture massive des données jurisprudentielles à une échéance proche a provoqué un climat d'émulation et suscité de multiples questions dont se sont saisies la DILA et l'association Open Law, afin d'impulser dès les origines une dynamique collaborative dans la construction des futurs modèles de diffusion des arrêts.

Les résultats du programme Open "Case" Law sont formalisés dans ce Livre Blanc, officiellement présenté à la DILA le 15 décembre.

Besoins, pistes de solution et points de vigilances y sont inventoriés afin de faciliter l'écriture d'une feuille de route pour la poursuite des travaux. Ce document est un premier socle de réflexion à partir duquel les acteurs intéressés pourront envisager un approfondissement de l'analyse des enjeux et étudier les moyens de l'ouverture de la jurisprudence.

1. <http://www.dila.premier-ministre.gouv.fr/activites/experimentations/open-case-law-2016>

2. <http://www.dila.premier-ministre.gouv.fr/activites/experimentations/les-programmes-d-innovation-open-law-le-droit-ouvert>

Préface

Depuis 2014, le mouvement « Open Law » consolide une communauté réunissant toute personne ou institution désirant contribuer à ce mouvement en faveur d'une plus grande et d'une meilleure diffusion du droit.

Les premiers programmes portés par l'association Open Law* le Droit Ouvert ont permis de mettre en valeur des projets qui ont pu continuer à se développer ensuite. Ainsi, le projet data.asso, distingué dans le cadre du concours co-organisé par Open law et la DILA sur la thématique de l'accès au droit, a été retenu dans le cadre d'un Programme d'investissement d'avenir (PIA) et a bénéficié d'un accompagnement par le ministère de la Jeunesse et des Sports.

Le programme Open Law Europa a permis en 2015 de franchir une nouvelle étape majeure avec le lancement de « communs », ces socles de connaissances partageables par l'ensemble de la communauté juridique.

Le présent livre blanc est consacré au cinquième programme d'Open Law. Ce programme s'est concentré sur l'open data des décisions de jurisprudence du fond (premier et second degrés) inscrit dans la loi pour une République numérique.

Ce cinquième programme, très riche, comme l'illustre ce livre blanc, a permis de travailler sur les enjeux de l'anonymisation préalable des décisions, leur enrichissement sémantique et leur diffusion.

Le hackathon réalisé lors de ce programme a favorisé la création de nouveaux services autour de deux thématiques importantes : la création industrielle des identifiants européens de la jurisprudence (ECLI) et la création de services innovants pour le ministère de la Justice.

Il a été un moment d'émulation associant secteurs public et privé, acteurs de toutes tailles, afin de faire progresser la diffusion publique des données de la justice.

La DILA est particulièrement fière d'y avoir été associée aux côtés du ministère de la Justice, de la Cour de cassation, de la Cour des comptes, du Conseil d'État, du Conseil constitutionnel, de la CNIL, de l'INPI, d'Etalab, et de l'Ordre des avocats de Paris.

Notre participation à ce programme confirme notre engagement depuis plusieurs années maintenant en faveur de l'ouverture des données et permet de réaffirmer que l'open data est un facteur d'égalité d'accès aux données.

Ce livre blanc est la concrétisation des avancées qui se sont dégagées tout au long du programme et que nous vous invitons à découvrir au fil des pages.

Véronique Lehideux

Directrice adjointe de l'information légale et administrative

**Le mot de la
coordinatrice
du programme**

“Ces travaux ont été une formidable occasion de réunir producteurs de jurisprudence, utilisateurs et réutilisateurs pour partager autour de leurs problématiques respectives.

Ouvrir les données juridiques, et les données de jurisprudence en particulier, c’est ouvrir un immense terrain de jeu pour l’application des techniques les plus avancées de traitement des données. Pour l’intérêt final du justiciable et de la justice elle-même : espérance de gain de cause et de dédommagement, pilotage statistique d’activité de tribunaux ou de résultats. Et par intérêt commercial. Car l’open data c’est l’opportunité de développer des produits et services commerciaux innovants et disruptifs.

L’utilisation des méthodes de travail de co-construction et d’innovation ouverte par l’ensemble des acteurs producteurs, utilisateurs et LegalTech est la clé de la réussite de ces initiatives, combinaisons d’inventivité et de rappels aux fondamentaux du droit.

Gageons que la démarche décrite dans ce Livre Blanc trace la voie pour relever ensemble les défis de la normalisation ab initio des formats de rédaction et de description des textes juridiques. Pour que les principes et méthodes de la légistique s’étendent et s’appliquent également à la jurisprudence.

Je tiens à souligner particulièrement le travail et l’implication de Sumi Saint-Auguste (Éditions Lefebvre Sarrut) et Camille Charles (association OpenLaw) sans qui ce Livre Blanc n’aurait pu voir le jour”.

Guiraudé Lame

Administratrice Open Law

Coordinatrice scientifique du programme Open Case Law

L'association Open Law

Fondée en 2015, l'association Open Law*, le Droit Ouvert est un espace de travail et d'expérimentation pour l'ensemble des acteurs du monde du droit qui innovent dans un modèle collaboratif et ouvert. Cette réunion permet l'identification de convergences entre les besoins qui guident l'orientation de nos travaux. Nos programmes de co-crédation numérigue ont vocation à déboucher sur des communs (recommandations : référentiel, livre blanc, modèles de statut, cahier des charges, etc. ou briques techniques : entrepôt de données, modèle, etc.) qui, à long terme, donneront à tous ceux qui les consultent ou les utilisent les moyens d'outiller leur réflexion et de prototyper de nouveaux services.

Le rôle d'Open Law est donc de servir une communauté qui œuvre collectivement à la transformation numérigue du monde du droit.

Innovation, ouverture et collaboration résument les méthodologies de travail de l'association et caractérisent l'ensemble de ses livrables.

The logo for Open Law* is a red square containing the text "OPEN LAW*" in white, bold, sans-serif capital letters. The asterisk is slightly smaller and positioned to the right of the word "LAW".

**OPEN
LAW***

* Le droit ouvert



À noter

Ce document est interactif, vous pouvez cliquer sur les liens du sommaire ainsi que sur ceux des sites internet afin d'être redirigé.



Sommaire

Préface	6
Le mot de la coordinatrice du programme	8
L'association Open Law	10

1 Le programme Open Case Law 14

Contexte légal	16
Objectifs du programme	19
Les partenaires du programme	20
Paroles de producteurs	20
Parole de réutilisateurs	21
Déroulé du programme	23

2 Le circuit de diffusion des décisions de justice 28

Les données et les flux	30
Les acteurs et les fonctions	36

3 Les enjeux de l'open data jurisprudentiel 42

Intégrer le service public de la donnée	45
Identifier le périmètre de l'open data jurisprudentiel	47
Protéger la vie privée du justiciable	51
Garantir la qualité des données et leur facilité de réutilisation	61
Bouleversement des usages et des business models	67
Conclusion	70
Synthèse	72

1

**Le programme
Open Case Law**

Contexte légal

Depuis plusieurs années, la France est engagée dans une dynamique d'ouverture des données détenues par les acteurs publics (open data). Si les décisions de justice ne sont pas des «données publiques» au sens de la loi CADA, elles ont bénéficié de ce mouvement général d'ouverture.

Une qualification complexe de la donnée jurisprudentielle³

L'accès aux décisions de justice est conditionné aujourd'hui par le caractère public de ces décisions. En vertu du principe de l'indépendance de la justice, les jugements, ordonnances, décisions ou arrêts ne sont en effet pas considérés comme des documents administratifs et donc hors du champ de la loi CADA⁴.

La nature des décisions de justice est ainsi précisée dans un avis rendu par la Commission d'Accès aux Documents Administratifs en 2010 : «(...) ne constituent pas des documents administratifs communicables en vertu de l'article 2 de la loi du 17 juillet 1978 (CE, 7 mai 2010, Bertin). Toutefois, il résulte de l'article 11-3 de la loi n° 72-626 du 5 juillet 1972 que les tiers sont en droit de se faire délivrer copie des jugements prononcés publiquement. La commission en déduit que l'accès à ces jugements constitue un droit pour toute personne et que ces derniers sont donc constitués d'informations publiques au sens de l'article 10 de la loi du 17 juillet 1978.»

Même si la décision est rendue au nom du peuple français, il n'en demeure pas moins que les juges se doivent de juger sur les faits pour dire le droit, et que ces derniers comportent des «données à caractère personnel».

En somme, les décisions de justice sont composées de données à caractère personnel (notamment en ce qui concerne les faits d'espèce) et de données publiques en ce qui concerne le droit.

L'accès en ligne aux décisions de justice est ainsi prévu par les textes depuis le décret du 7 août 2002 relatif au service public de la diffusion du droit par

3. Mémoire de Bertrand Cassar sous la direction de Thomas Saint-Aubin (Université Paris I), [http://openlaw.fr/images/7/7d/Anom_Pseudo_Donn%C3%A9es_Justice_\(CASSAR_B\).pdf](http://openlaw.fr/images/7/7d/Anom_Pseudo_Donn%C3%A9es_Justice_(CASSAR_B).pdf)

4. Loi n. 78-753 du 17 juillet 1978

l'internet⁵, prévoyant la diffusion gratuite des données juridiques sur Legifrance. **Les praticiens du droit et les justiciables peuvent aujourd'hui consulter la quasi-totalité des arrêts des Cours suprêmes**, ainsi qu'une sélection des arrêts de Cour d'appel et cours administratives d'appel (pour le premier degré, la proportion de jugements de tribunaux accessibles sur Legifrance reste pour l'instant infime). **Depuis septembre 2015, ces jeux de données ont été ouverts et sont désormais disponibles sur data.gouv.fr⁶ en open data et susceptibles d'être réutilisés⁷.**

La discussion, puis l'adoption de la Loi pour une République numérique⁸ du 7 octobre 2016 a encore accentué cette politique en faveur de l'open data. S'agissant des données jurisprudentielles, ce texte consacre un véritable changement de paradigme. Ses articles 20 et 21 disposent en effet que la totalité des décisions de premier et second degré seront mises à disposition en open data sous réserve du principe du respect de la vie privée des personnes concernées et sans préjudice des dispositions particulières qui régissent l'accès aux décisions de justice de l'ordre judiciaire et à leur publicité.

Article 20

L'article L. 10 du code de justice administrative est complété par quatre alinéas ainsi rédigés :

« Ces jugements sont mis à la disposition du public à titre gratuit dans le respect de la vie privée des personnes concernées.

« Cette mise à disposition du public est précédée d'une analyse du risque de ré-identification des personnes.

« Les articles L. 321-1 à L. 326-1 du code des relations entre le public et l'administration sont également applicables à la réutilisation des informations publiques figurant dans ces jugements.

« Un décret en Conseil d'Etat fixe, pour les jugements de premier ressort, d'appel ou de cassation, les conditions d'application du présent article. »

5. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000413818&categorieLien=id>

6. <http://www.dila.premier-ministre.gouv.fr/actualites/actualites/open-data-la-dila-ouvre-4-nouveaux-jeux-de-donnees>

7. Il s'agit des bases CAPP, INCA, CASS et JADE.

8. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000033202746&dateTexte=20161107>

Article 21

Le chapitre unique du titre Ier du livre Ier du code de l'organisation judiciaire est complété par un article L. 111-13 ainsi rédigé :

« Art. L. 111-13. - Sans préjudice des dispositions particulières qui régissent l'accès aux décisions de justice et leur publicité, les décisions rendues par les juridictions judiciaires sont mises à la disposition du public à titre gratuit dans le respect de la vie privée des personnes concernées.

« Cette mise à disposition du public est précédée d'une analyse du risque de ré-identification des personnes.

« Les articles L. 321-1 à L. 326-1 du code des relations entre le public et l'administration sont également applicables à la réutilisation des informations publiques figurant dans ces décisions.

« Un décret en Conseil d'Etat fixe, pour les décisions de premier ressort, d'appel ou de cassation, les conditions d'application du présent article. »

Cette volonté d'ouverture s'est également exprimée dans le cadre du **Partenariat pour un Gouvernement Ouvert**, dont la France est membre depuis 2014 et qu'elle préside depuis le 26 septembre 2016 (pour une durée d'un an). Au titre de cette initiative, un "Plan d'action national pour une action publique transparente et collaborative (2015-2017)"⁹ a été adopté en 2015. Celui-ci est composé de 26 engagements. **Le douzième porte spécifiquement sur l'ouverture de la jurisprudence**¹⁰.

9. <https://www.etalab.gouv.fr/plan-daction-national>

10. <https://suivi-gouvernement-ouvert.etalab.gouv.fr/fr/Engagement12.html>

Objectifs

Ce programme vise à anticiper de manière collective les enjeux techniques et juridiques relatifs à l'ouverture de la jurisprudence du fond (cf. partie 3), afin d'être en mesure de produire une première lecture prospective de ce que pourrait être, demain, le service public de la donnée publique juridique.

Plus concrètement, le Programme 5 – Open « Case » Law s'est employé à faire émerger une vision concrète de la mise en œuvre de l'open data en matière de décisions de justice en :

- * établissant un état des lieux,
- * identifiant les grandes questions,
- * explorant des hypothèses pour la suite.

Les acteurs du programme mettent ainsi leur expérience de **tiers neutre** (Open Law) et de **diffuseur public** (DILA) à disposition du législateur afin de faciliter la rédaction des décrets d'application de la loi République numérique, garantissant une assise collaborative à la démarche.

Les **priorités du projet** ont été de cartographier rigoureusement le service de diffusion actuel des données jurisprudentielles, de se pencher sur les problématiques d'anonymisation, de développer une dynamique de collaboration avec le secteur privé et de faire remonter les attentes des consommateurs et utilisateurs des données publiques juridiques.



Les équipes du hackathon en plein travail, le 16 novembre 2016, au sein du Village de la Legal Tech/Paris Open Source Summit

Les partenaires du programme

Ont contribué activement, en tant que partenaires du programme : la mission ETALAB, le ministère de la Justice, la Cour de cassation, la Cour des comptes, le Conseil d'État, la CNIL, l'INPI et l'Ordre des avocats de Paris.

Nous remercions également l'ensemble des participants aux différents ateliers qui, par leur assiduité et la richesse de leurs interventions, ont considérablement enrichi la réflexion sur l'Open data jurisprudentiel.

Paroles de producteurs

✱ Le ministère de la Justice

« Le ministère de la Justice ne pouvait que s'investir avec enthousiasme dans une démarche visant à amener l'ensemble des acteurs de l'ouverture des décisions de justice – producteurs, utilisateurs et réutilisateurs des données – à réfléchir ensemble et de façon décloisonnée sur les perspectives de l'open data de la jurisprudence. Cette nouvelle exigence citoyenne sera le vecteur d'un renouvellement des services qui seront proposés aux justiciables comme aux professionnels de la justice. La qualité des échanges au sein du programme a permis d'enrichir les visions de chacun et de mieux partager les enjeux de la mise en œuvre de la loi du 7 octobre 2016 pour une République numérique. »

Eloi Buat-Ménard

*Adjoint à la sous-directrice de l'organisation judiciaire et de l'innovation
Direction des Services judiciaires, ministère de la Justice*

✱ La Cour de Cassation

« Les objectifs du programme étaient convergents avec les démarches déjà engagées par la Cour de cassation en faveur de l'open data des décisions de justice. Le programme a répondu aux attentes de la Cour, en permettant la

mise en œuvre d'échanges de qualité avec l'ensemble des acteurs intéressés et en encourageant les initiatives innovantes en faveur de la transparence de la justice et de la connaissance du droit. »

Ronan Guerlot

Conseiller référendaire, Adjoint du directeur

Cour de Cassation, Service de documentation études et du rapport

*** Le Conseil d'État**

« Le Conseil d'État, invité à participer au programme « Open Case Law », l'a suivi avec d'autant plus d'intérêt que les axes de réflexion choisis coïncidaient avec ceux que lui dictaient dans sa démarche de diffusion ses propres objectifs et les conséquences de l'adoption de la loi «pour une République numérique». Il a trouvé dans ce processus collectif l'opportunité de rencontrer des acteurs de la diffusion du droit qui étaient pour lui des interlocuteurs nouveaux et, à travers eux, un précieux complément à sa réflexion interne. »

Pierre-Yves Martinie

Responsable du service de diffusion de la jurisprudence

Conseil d'État

Paroles de réutilisateurs

*** Un représentant des professions juridiques : l'ordre du Barreau de Paris**

« L'Ordre des Avocats participe activement à toutes les actions améliorant l'accès au Droit pour le citoyen et le justiciable. Ce programme lui a permis de partager avec d'autres acteurs l'idée qu'une meilleure pratique du droit et de la défense judiciaire est favorisée par un libre accès de tous, à niveau égal, aux données publiques juridiques. Les représentants de l'Ordre ont apprécié la qualité des échanges des diverses réunions. Il envisage de poursuivre activement ce dialogue avec les acteurs de ce programme jusqu'à l'ouverture de ces nouveaux fonds sur les sites publics et au-delà, en y intégrant, à l'exemple de ses homologues belges, les promesses d'un usage optimisé assuré par l'intelligence artificielle. »

*** Un éditeur : Dalloz**

« Dalloz a participé au programme Open Law pour rencontrer les différents acteurs de l'ouverture des données juridiques et alimenter ses réflexions sur

l'intégration de Dalloz dans la dynamique de l'open access : jusqu'où peut-on aller ? à quel rythme ? dans quelles conditions ? avec quels partenaires ? ... Open Law nous a également permis de présenter simplement à l'ensemble du marché la position de Dalloz sur le sujet et d'en discuter de façon ouverte et constructive. En complément, le Paris Open Source Summit et son village de la Legal Tech ont été l'occasion d'aller au bout de l'expérience en rencontrant, au-delà des institutions, tous les types d'utilisateurs finaux. »

Camille Szejnhorn

Directrice des marchés juridiques

Éditions Dalloz

✱ Une legal start-up : Case Law Analytics

Case Law Analytics s'emploie à mesurer et visualiser l'aléa judiciaire observé dans les décisions de justice.

« On comptait en 2012 à peine plus de magistrats qu'au début de la Restauration alors que la population française a plus que doublé en deux siècles » (Farcy, 2015). Or l'intelligence artificielle permet désormais, en utilisant les données de disposer de probabilités très fiables sur l'issue de certains contentieux de masse pour y favoriser les modes amiables. »

Jacques Levy-Vehel

Inria, co-fondateur de Case Law Analytics

Trois ateliers ont été mis en place pour réfléchir collectivement sur les problématiques suivantes :

- ✱ Anonymisation/dépersonnalisation,
- ✱ European Case Law Identifier (ECLI),
- ✱ Service public de la donnée publique juridique.

Déroulé du programme

Les travaux réalisés dans ce cadre ont permis en outre de fournir la matière première nécessaire à l'organisation d'un hackathon qui s'est déroulé le 16 novembre au Paris Open Source Summit¹¹.

✱ Atelier 1 : État de l'art des solutions et pratiques d'anonymisation et de la dépersonnalisation des décisions de justice

Date : 29 septembre 2016

L'atelier a démarré par une présentation détaillée par la CNIL des enjeux liés à la diffusion des décisions de justice et de l'utilité d'un potentiel «pack de conformité» à destination des producteurs de données juridiques (s'agissant de la dépersonnalisation de leurs décisions). La trentaine de participants s'est ensuite répartie en deux sous-groupes.

Le premier a réfléchi aux champs essentiels d'un référentiel des métadonnées pertinentes pour une décision de justice et présentes dans les données. Le second s'est intéressé à la chaîne de risque, en partant du constat d'un impératif de transparence et de clarté en matière de responsabilité.

✱ Atelier 2 : Les normes. Quel ECLI pour les décisions du fond ?

Date : 5 octobre 2016

Après des rappels théoriques sur la genèse d'ECLI et les orientations prises dans quelques États-membres de l'Union européenne, la DILA a présenté son projet de référentiel des juridictions, qui entre dans la grammaire d'écriture d'ECLI.

L'atelier s'est ensuite scindé en trois sous-groupes qui ont travaillé sur les thématiques suivantes :

- ✱ propositions et discussion autour du référentiel de la DILA,
- ✱ enrichissement des décisions par des métadonnées,
- ✱ chaîne de valeur autour du référencement ECLI, en préparation du hackathon du 16 novembre.

11. <https://www.meetup.com/fr-FR/Legal-Innovation-Paris/events/233816366/>

12. Les comptes-rendus détaillés de ces ateliers peuvent être fournis sur simple demande adressée à contact@openlaw.fr

✱ Fab Law : co-construction du service public de la donnée juridique

Date : 25 octobre 2016

Cet atelier a permis de donner la parole à l'ensemble des acteurs du circuit de diffusion des décisions de jurisprudence, du producteur à l'utilisateur final de la donnée (cf. Tableau de synthèse).

Après une brève présentation de leurs bases par le Conseil d'État, la Cour de cassation et la Cour des comptes, ainsi qu'une présentation des missions de diffusion de la DILA, un représentant d'Etalab a exposé les premiers résultats de la consultation publique alors en cours sur le futur Service public de la donnée¹³.

Les réutilisateurs de données juridiques (éditeurs juridiques et legal start-ups) ont finalement présenté leurs usages de ces données, ainsi que les attentes et les craintes associées à leur ouverture prochaine en open data.

L'atelier s'est ensuite scindé en deux.

Le premier sous-groupe s'est inspiré du canevas de la consultation sur la mise en œuvre du Service public de la donnée conduite par Etalab, pour faire valoir les besoins propres à la sphère juridique. L'ensemble des participants s'est accordé sur l'importance de garantir une fraîcheur optimale des données et leur diffusion dans un format interopérable (cf. couverture intérieure en fin de livret).

Le second sous-groupe a traité la question des circuits de diffusion des données juridiques. Il ressort de leur discussion la nécessité de prévoir une granularité suffisante pour prendre en compte les besoins propres à chaque catégorie d'utilisateurs (cf. couverture intérieure en début de livret).

Le hackathon

Dernière rencontre publique de ce programme, un hackathon s'est déroulé le 16 novembre dans le cadre du Paris Open Source Summit¹⁴, avec pour objectif de relever deux défis :

- ✱ Le déploiement de la norme ECLI pour les juridictions du fond, présenté par la DILA, dans un double objectif d'interopérabilité et de création d'opportunités de service à valeur ajoutée,
- ✱ La proposition de services innovants puisant dans l'open data jurisprudentiel ainsi que dans une sélection de données statistiques,

.....
13. <https://www.etalab.gouv.fr/consultation-spd>

fournies par le ministère de la Justice, susceptibles d'être utiles à ce dernier dans sa mission d'administration des tribunaux.

Cinq projets issus d'équipes mixtes (acteurs publics et privés) ont proposé à l'issue de la journée de travail :

- ★ **Réalisation d'un prototype de datavisualisation des indemnités allouées par les cours d'appel.** Grâce à un outil de machine learning et un algorithme de compréhension du langage juridique, il a été possible de déterminer des probabilités d'obtenir des indemnités, selon des tranches à la granularité très fine, quel que soit le sujet envisagé. Cet outil, potentiellement utile pour les magistrats et la Chancellerie, résulte d'une collaboration entre la startup Predictice et la R&D du groupe Éditions Lefebvre Sarrut ;
- ★ **Définition d'une méthode pour créer automatiquement des ECLI sur les décisions de jurisprudence.** Une solution est proposée à partir des flux XML reçus en utilisant des référentiels et un outil de traitement automatique du langage (TAL) afin de sémantiser et donner une URI à chaque décision de jurisprudence de manière descendante depuis la Cour de cassation vers les cours d'appels et les tribunaux. Chaque décision de justice sera enrichie et accessible avec un moteur de recherche puissant. Le prototype permet grâce aux métadonnées recensées et enrichies d'obtenir un corpus de décisions facilement exploitable. Ce prototype a été développé par Mondeca et le Groupe Revue Fiduciaire ;
- ★ **Outil de gestion et d'étude du positionnement des brevets.** Le prototype vérifie dans la base de jurisprudence de l'INPI les données concernant des brevets similaires et les citations de décisions de jurisprudence provenant de la doctrine contenue dans Lamyline y afférant. Cet outil proposé par Wolters Kluwer France et Questel facilite le dépôt de brevets et le traitement des litiges pouvant survenir ;
- ★ **Outil d'aide à l'évaluation du risque de ré-identification sur les décisions de jurisprudence déjà anonymisées.** L'enjeu est de pouvoir sécuriser le processus d'anonymisation. Le prototype permet grâce à l'application de critères d'identification d'analyser une décision de justice et de vérifier si le résultat permet ou non la ré-identification des parties et de mesurer ainsi le degré d'anonymisation. Cet outil a été élaboré par la startup JurisLegaTech ;

- ★ **Outil de quantification du montant de la pension alimentaire.** Le prototype grâce à l'apport des technologies prédictives permet de calculer en fonction de plusieurs critères le montant estimé des pensions. Chaque montant est pondéré d'une probabilité. Cet outil a été développé par la startup Case Law Analytics.

Le cercle vertueux de l'innovation ouverte

La mobilisation de la communauté Open Law a permis de rassembler autour de la table l'ensemble des acteurs pertinents. Acteurs publics de la diffusion du droit et des données (DILA, Etalab, ministère de la Justice, CNIL), Cours suprêmes, legal start-up, éditeurs et professions réglementées ont ainsi pu échanger sur un mode ouvert à propos de leurs attentes, contraintes et de leurs projets en cours concernant la diffusion des données jurisprudentielles. Le Hackathon a permis d'expérimenter sous des formes inédites à partir de jeux de données juridiques publiques des nouveaux services à destination des acteurs publics comme des justiciables.

La multiplication de ces temps d'échanges a déjà généré de multiples bénéfices : éclosion de relations stratégiques entre secteur public et privé au service de l'accélération de l'ouverture, démythification des positionnements réciproques des acteurs impliqués et accélération des transferts de savoir et de compétence.

Préconisation #1

Poursuivre ces rencontres entre acteurs publics et privés dans le cadre d'un groupe de travail pérenne consacré à l'open data jurisprudentiel. Une piste de travail pourrait être au démarrage de se concentrer sur une juridiction ou un corpus donné de décisions du fond pour travailler collectivement à partir de cas d'usage concrets (approche bottom-up).

**Avertissement : cette section
doit se lire en se reportant
à l'infographie reproduite
sur la couverture intérieure
gauche de ce livret.**

2

**Le circuit de
diffusion des
décisions de
justice**

Les données et les flux

L'ordre judiciaire

L'ordre judiciaire comprend des juridictions du fond, de première (TGI et TI) et seconde instance (cour d'appels), ainsi que de la Cour de cassation. Ces juridictions rendent des décisions de deux types : civiles et pénales.

Actuellement, le service de documentation, des études et du rapport (SDER) de la Cour de cassation tient **deux bases de données de jurisprudence**, ainsi que le prévoit l'article R. 433-3 du code de l'organisation judiciaire¹⁵ :

- * **La base dite « Jurinet »** qui comprend l'ensemble des décisions de la Cour de cassation, ainsi qu'une sélection de décisions des juridictions du fond présentant un intérêt particulier ;
- * **La base dite « JuriCA »** qui comprend l'ensemble des décisions civiles motivées des cours d'appel. La base JuriCA a permis de collecter environ 150 000 arrêts civils par an depuis 10 ans, soit un stock d'environ 1,5 million de décisions. A l'avenir, le volume de décisions pouvant être collectées augmentera considérablement (voir l'encadré ci-dessous).

Actuellement, deux flux de données sont accessibles aux réutilisateurs¹⁶ :

- * **En diffusion publique** : les arrêts contenus dans la base Jurinet sont diffusés dans les conditions applicables au service public de diffusion du droit par l'internet, comme le prévoit l'article R. 433-3 du code de l'organisation judiciaire ; cela concerne les décisions « publiées » de la Cour de cassation (base CASS), les décisions dites « inédites » de la Cour de cassation (base INCA) et la sélection de décision des juridictions du fond (base CAPP) ; ces décisions sont diffusées, après pseudoanonymisation, sur le site Légifrance et en open data sur data.gouv.fr¹⁷.

15. <https://www.legifrance.gouv.fr/affichCodeArticle.do?cidTexte=LEGITEXT000006071164&idArticle=LEGIARTI000018921800>

16. Les magistrats disposent d'un accès dédié à ces bases par l'intermédiaire du RPVJ (réseau privé virtuel justice)

17. Le flux « Légifrance » et le flux « data.gouv.fr » concernent strictement les mêmes décisions (celles qui sont contenues également sur Jurinet)

- ★ **En diffusion restreinte** : les arrêts contenus dans la base JuriCA sont accessibles dans les conditions fixées par la CNIL en 2012¹⁸ : soit par l'intermédiaire d'un abonnement au fonds de concours de la Cour de cassation (avec livraison quotidienne), soit dans le cadre de conventions de mise à disposition gratuite de certaines décisions utiles à des travaux de recherche ; conformément aux règles prescrites par la CNIL, ces décisions sont transmises non anonymisées, à charge pour l'utilisateur de les anonymiser avant toute diffusion publique et de respecter plus généralement les obligations de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

Un flux pseudonymisé, dans lequel seules les métadonnées contiennent encore le nom des parties, était également diffusé sous licence et avec convention de services ; la DILA a mis fin aux conventions de services fin 2016 et arrêtera ce flux courant 2017.

A ce jour, l'anonymisation des décisions diffusées sur Légifrance et data.gouv.fr est sous-traitée à la Dila qui utilise comme certains éditeurs privés la technologie Temis (Expert System) et ses cartouches de fouille de texte. Son modèle est décrit sur le site : <https://www.legifrance.gouv.fr/Bases-de-donnees/Contenus/Jurisprudence-judiciaire>. La Cour de cassation a engagé un processus d'internalisation de l'anonymisation.

Impacts de la loi Lemaire : le point de vue de la Cour de cassation

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). →

18. Délibération n°2012-245 du 19 juillet 2012 concernant la base de jurisprudence Jurinet (<https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000026308530>) et délibération n°2012-246 du 19 juillet 2012 concernant la base de jurisprudence JuriCA (<https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000026308531>)

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 décisions du juge des libertés et de la détention).

À 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. »

L'ordre administratif

L'ordre administratif est composé, pour les juridictions non spécialisées, des tribunaux administratifs (TA), des cours d'appel administratives (CAA) et du Conseil d'État (CE). Le centre de recherches et de diffusion juridiques du Conseil d'État est également en charge de la diffusion des décisions du Tribunal des conflits, qu'assure parallèlement la Cour de cassation.

La volumétrie des décisions administratives est sans commune mesure avec le judiciaire. En effet, on dénombre quelque 230 000 décisions rendues chaque année.

La juridiction administrative alimente quatre bases de décisions : deux exclusivement internes, **Ariane** et **Ariane archives**, et deux d'accès public : **Jade**, diffusée par la DILA sur Legifrance et référencée en open data sur data.gouv.fr ; **Ariane web** sur le site internet du Conseil d'État.

On peut rappeler utilement que la collecte et la diffusion des décisions du Conseil d'État trouve son origine dans le Recueil Lebon ; il s'est agi dès le départ de proposer un corpus de décisions dûment classé et indexé : le Recueil est constitué de décisions dans leur intégralité et en complément d'une large sélection de décisions simplement titrées et indexées dans un plan de classement avec une granularité fine, ce qui est a priori contradictoire avec les impératifs d'exhaustivité inhérents à l'open data.

La démarche principale du Conseil d'État est toujours axée sur la diffusion de la jurisprudence (au sens large), c'est-à-dire qu'elle s'exerce sur des sélections de décisions qui sont systématiquement hiérarchisées selon leur apport jurisprudentiel et, le cas échéant, assorties d'une analyse de cet apport. Ainsi, sont écartées d'emblée les décisions qui ne présentent pas d'intérêt jurisprudentiel : ordonnances autres que celles de référé rejetant une requête pour irrecevabilité manifeste non régularisable ou donnant acte d'un désistement, décisions non motivées de non-admission des pourvois en cassation, etc.

La base Ariane qui, étant à usage strictement interne, n'est pas anonymisée, contient ainsi :

- ★ 160 000 fichiers de décisions du Conseil d'État, 43 000 fichiers d'analyse associés,
- ★ 300 000 arrêts de CAA et 5000 analyses,
- ★ 19 000 jugements de TA qui correspondent aussi à une sélection par les tribunaux eux-mêmes, et 6000 analyses,

- * les décisions du Tribunal des conflits et leurs analyses,
- * plusieurs dizaines de milliers des conclusions des commissaires du gouvernement ou rapporteurs publics.

Cette base interne Ariane est complétée par une base d'archives, Ariane archives, également non anonymisée et qui donne accès à la quasi-intégralité des documents produits par les TA et les CAA, inclus les jugements et décisions sans valeur jurisprudentielle.

Ariane Web, accessible depuis 2011 sur le site internet du Conseil d'Etat, a été conçue comme une extraction anonymisée de la partie d'Ariane (interne) correspondant à la jurisprudence au sens strict du Conseil d'Etat, des CAA et du tribunal des conflits

« Cette base de jurisprudence donne accès à plus de 230 000 documents : décisions et avis contentieux du Conseil d'État et des cours administratives d'appel, analyses de ces décisions et avis ayant été retenus pour leur apport à la jurisprudence, ainsi qu'une sélection de conclusions de rapporteurs publics. Ces documents sont présentés dans des fonds distincts ».

Enfin, la base Jade, qui est la seule base de jurisprudence administrative sur laquelle puissent être souscrites des licences de réutilisation, offre des fonds de décisions équivalents, pour le Conseil d'Etat et le Tribunal des conflits, à ceux d'Ariane Web et d'Ariane interne, supérieurs, pour les CAA à ceux d'Ariane Web, ce qui en fait à cet égard la base publique la plus complète pour la juridiction administrative.

Actuellement, trois flux de données sont accessibles aux réutilisateurs :

- * **le flux « fonds de concours »** : décisions de l'ordre administratif, des CAA jusqu'au Conseil d'Etat, ainsi que le Tribunal des conflits (les TA sont à venir prochainement) ; les données sont livrées anonymisées, sur abonnement avec redevance (catalogue de tarifs sur des flux thématiques, ce qui implique leur tagguage thématique très en amont) ; livraisons quotidiennes à partir du jour qui suit la date de lecture jusqu'à deux semaines,
- * **le flux Legifrance (JADE)** : les décisions du Conseil d'Etat à l'exception des ordonnances autres que celles de référé et décisions de non-admission des pourvois en cassation, une très

large sélection d'arrêts des CAA (offre plus large donc que celle qui est publiée sur Arianeweb, le site du Conseil d'État) et le Tribunal des conflits ; l'anonymisation est opérée en amont par le Conseil d'État et la DILA la finalise pour diffusion sur Legifrance et data.gouv,

- * **Le flux en open data sur data.gouv.fr** : même sélection que pour Legifrance avec même périodicité et même anonymisation.

L'anonymisation des décisions du Conseil d'État est assurée par celui-ci au moyen d'un logiciel réalisé par la société AlterSystems et intégré à la chaîne de versement des décisions. Il répond à des spécifications légèrement différentes de celles du judiciaire (Legifrance) pour la substitution des éléments à anonymiser (A et B au lieu de X, Y...).

Impacts de la loi République numérique : le point de vue du Conseil d'État

Pour les décisions des juridictions administratives, l'enjeu principal face à l'obligation de mise à disposition en open data reste l'augmentation conséquente de la masse de décisions à traiter (et anonymiser). On passe d'une logique de sélection sous un angle jurisprudentiel à une exigence d'exhaustivité qui permet des explorations algorithmiques ou d'autres alors empruntés aux technologies de big data.

(cependant qu'une sélection seulement est proposée sur le site ArianeWeb), le Conseil d'État se met en ordre de marche pour préparer le basculement en open data par défaut. Le défi majeur est celui de l'anonymisation car le processus actuel, qui est semi-automatisé et nécessite un contrôle humain attentif et assidu, semble difficile à transposer lorsqu'il s'agira de traiter de plus gros volumes. Le Conseil d'État place ses espoirs dans des technologies plus perfectionnées et plus fiables pour y remédier et y travaille dès à présent.

Les acteurs et les fonctions

- * **cours suprêmes** (Conseil d'État et Cour de cassation),
- * **cours d'appel** (CAA et CA),
- * **juridictions judiciaires de premier degré** (TGI, TI, juge de proximité),
- * **juridictions administratives du premier degré** (TA),
- * **organismes et autorités publiques** (INPI, Autorité de la concurrence, etc),
- * **universitaires** (étudiants et chercheurs),
- * **justiciables** (personnes physiques, personnes morales),
- * **professionnels du droit** (juristes dont CPI, professions réglementées, magistrats),
- * **prestataires de services** (start-up et acteurs plus anciens de type éditeurs et grands cabinets d'avocats ayant développé des services reposant sur la réutilisation / l'enrichissement / l'accès aux données de jurisprudence),
- * **société civile et tiers secteur** (presse généraliste, associations de défense des droits de l'homme, associations de consommateurs, collectifs citoyens, etc.).

Les producteurs de données

Les producteurs de données sont les magistrats qui rédigent les décisions de justice (au sein des juridictions administratives et judiciaires), mais également l'ensemble des personnes qui participent à leur anonymisation et leur enrichissement (métadonnées, commentaires, résumés, titrages).

Dans la situation actuelle, les éditeurs privés, ou des organismes publics comme l'INPI, qui proposent des bases de données enrichies, voire les prestataires qui développent des solutions d'anonymisation, pourraient être rapprochés de cette catégorie des producteurs.

Le fait déclencheur

Le fait déclencheur de la publication est le prononcé de la décision de justice, qui lui confère son caractère « public »¹⁹. L'idée que seules les décisions juridiquement « définitives » puissent être systématiquement diffusées, un moment retenue par le Parlement, a été abandonnée comme ne correspondant pas à l'état du droit – les décisions mêmes non définitives ont déjà un apport en droit – et soulevant de très épineuses difficultés d'application.

Les diffuseurs

Les diffuseurs sont les entités qui délivrent la donnée auprès des usagers, professionnels du droit ou citoyens et entreprises, soit en permettant sa simple consultation (associée à des moteurs de recherche et des services plus ou moins sophistiqués), soit à travers sa mise à disposition en open data (et donc la possibilité d'extraire la donnée et de l'exploiter).

Aujourd'hui, on compte de multiples diffuseurs en France, entités publiques mais également privées, proposant des modèles d'accès gratuits ou payants :

- ★ **La DILA** : accès en libre consultation d'une sélection des décisions de justice sur Légifrance (bases CASS, CAPP, INCA et JADE) ; mise à disposition de ces bases en open data par référencement sur le site data.gouv.fr,
- ★ **ETALAB** : mise à disposition gratuite d'une sélection des décisions sur data.gouv.fr,
- ★ **Les Cours suprêmes** : le Conseil d'État propose sur son site internet un moteur de recherche permettant d'interroger sa base de données (Ariane Web). La Cour de cassation diffuse certains de ses arrêts, parfois accompagnés de notes explicatives, et l'ensemble de ses publications, sur son site internet. Actuellement, aucun moteur de recherche ne permet d'accéder aux bases CAPP, CASS et INCA sur le site de la Cour de cassation,
- ★ **Les organismes publics – et notamment l'INPI** : L'Inpi enrichit par des métadonnées les décisions rendues en matière de propriété industrielle, extraites de Jurica et recueillies

19. Cf Partie 3, « Identifier le périmètre de l'open data jurisprudentiel ».

auprès du TGI de Paris, proposant ainsi à la diffusion la quasi exhaustivité de ce contentieux spécifique. Anonymisées, les décisions sont en accès libre et gratuit sur le site de l'Institut, et disponibles à la réutilisation par le biais d'une licence Open data,

- * **Les éditeurs privés** : les éditeurs privés accèdent à un grand nombre de données auprès des juridictions, par le biais d'abonnements à des fonds de concours soumis à redevance²⁰. Ces données ne sont pas nécessairement anonymisées, leur accès étant encadré, pour les décisions des juridictions judiciaires, par les délibérations de la CNIL du 19 juillet 2012²¹ (données reçues non anonymisées, mais avec des obligations d'anonymisation et un contrôle de la Cour de cassation). Les données du Conseil d'État la fois sont fournies à la fois anonymisées (word) et anonymisables (XML).

Généralement, leur offre de service est fondée sur une offre d'accès payante à leurs bases et à des fonctionnalités de recherche approfondie à destination des professionnels du droit. Néanmoins, les bouleversements introduits par la loi République numérique commencent déjà à modifier le paysage économique, puisque certains éditeurs proposent désormais des accès gratuits à une sélection de données.

Chacun à leur niveau, ces diffuseurs proposent un certain nombre de services associés à l'exploitation des bases de données : moteurs de recherche, gestion des droits d'opposition, à la rectification et à l'oubli, archivage, etc.

.....

20. Le fonds de concours de la Cour de cassation permet de recevoir sur abonnement les données de JuriCA et de Jurinet.

21. Délibération n°2012-245 du 19 juillet 2012 concernant la base de jurisprudence Jurinet (<https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000026308530>) et délibération n°2012-246 du 19 juillet 2012 concernant la base de jurisprudence JuriCA (<https://www.legifrance.gouv.fr/affichCnil.do?id=CNILTEXT000026308531>)

Les (ré)utilisateurs

Les typologies (ré)utilisateurs sont aussi nombreuses que les usages envisagés pour ces données. Schématiquement, il est possible de les répartir en trois grandes catégories :

- * **usages à des fins éditoriales et commerciales** : les données alimentent l'offre de service d'acteurs économiques. Traditionnellement utilisées par les éditeurs juridiques, ces données sont désormais exploitées par de jeunes acteurs pour leur potentiel en termes d'analyse prédictive et de gestion du risque juridique,
- * **usages à des fins scientifiques** : ici, les données sont utilisées pour alimenter les travaux de recherche (histoire du droit, sociologie du droit, etc.) ou à des fins statistiques,
- * **usages à des fins informatives** : les données sont consultées pour la valeur informative de leur contenu, soit par les professionnels du droit dans l'exercice de leur métier, soit par la société civile en tant que protectrice de l'indépendance et l'impartialité de la justice, soit par le justiciable qui cherche à connaître le Droit.

La loi Lemaire - Quels impacts sur les circuits de diffusion ?

L'open data doit être envisagé comme une véritable opportunité de repenser la gouvernance de l'ensemble de la chaîne de production des données jurisprudentielles.

Il est nécessaire d'avoir une vision globale du cycle de vie de la donnée (de sa production à son archivage) qui intègre les enjeux liés à l'impératif de protection de la vie privée, mais également l'exercice potentiel de leur droit de rectification, leur droit de retrait et leur droit à l'oubli par les justiciables et les professionnels du droit.

Au-delà, la nécessité de prendre en compte les besoins des ré-utilisateurs, notamment s'agissant de l'intégrité des décisions de justice, justifie qu'elles circulent sous des régimes différents.

C'est donc le rôle de l'ensemble des acteurs qui doit être repensé, ainsi que leur position respective dans la chaîne de responsabilité liée aux dommages éventuels causés par la ré-identification après diffusion.

En résumé ci-après, un tableau synthétique permet d'identifier aisément le positionnement actuel de chaque acteur par rapport à la donnée jurisprudentielle.

En tant qu'acteurs	Je produis	Je délivre			J'utilise		Quel usage de la donnée ?
	Émission	en Open Data	en Open Access	en accès payant	Consultation	Exploitation	
Cours Suprêmes	OUI	NON	OUI	OUI	OUI	NON	Information et veille juridique pour exercice de leur profession;
Les Juridictions du fonds	OUI	NON	OUI*	OUI*			Certain gros cabinets d'avocats sont en capacité d'exploiter les base de données pour développer des services innovants.
Professionnels du Droit	NON	NON	NON	NON	OUI	NON	Publication en tant qu'opérateur de legifrance.fr et en tant qu'opérateur de Légifrance.
DILA	NON	OUI	OUI	NON**	NON	NON	Mise à disposition en tant qu'opérateur de data.gouv.fr
Etalab	NON	OUI	NON	NON	NON	NON	Publication, enrichissement
Éditeurs	NON	NON	NON	OUI	OUI	OUI	Exploitation dans le cadre de prestations de services à forte valeur ajoutée Usage type : prédictibilité, machine learning, analyse et mesure des risques, etc
Acteurs Legaltech	NON	NON	OUI	OUI	NON	OUI	Information juridique pour nourrir leurs projets de recherches, exploitation des données à des fins scientifiques (statistiques, modélisations etc)
Universitaires	NON	NON	NON	NON	OUI	OUI	Information juridique et accès au droit
Justiciables	NON	NON	NON	NON	OUI	NON	Information juridique pour exercer une forme de "surveillance citoyenne", s'assurer de l'indépendance et de l'impartialité de la Justice
Société civile et tiers secteurs	NON	NON	NON	NON	OUI	OUI	

* Une partie des arrêts de cours d'appel et et la totalité des arrêts de cours administratives d'appel sont accessibles sur Légifrance et sur data.gouv.

** La diffusion est gratuite, mais elle s'accompagnait, pour certains réutilisateurs, de conventions de services payantes qui ont pris fin au 31 décembre 2016

Avertissement : sont volontairement exclues de cette analyse les décisions du tribunal des conflits, des juridictions financières et du Conseil constitutionnel.

3

**Les enjeux de
l'open data
jurisprudentiel**

Les enjeux qui se dessinent sont d'ordre logistique et technique d'une part, et d'ordre sociétal et économique d'autre part.

Le principe d'open data par défaut augmente considérablement la volumétrie des données concernées, ce qui constitue une gageure puisqu'il s'agira de garantir une qualité optimale égale pour ce périmètre étendu des données de jurisprudence.

L'adoption des standards et bonnes pratiques de l'open data (format numérique et interopérabilité, identifiant ECLI) suppose que les chaînes de traitement adéquates soient mises en place, et que les règles de génération d'un ECLI soient posées.

Intégrer le service public de la donnée

La mission ETALAB a conduit entre le 29 septembre et le 20 octobre 2016 une consultation²² consacrée **aux six jeux de données de référence**, identifiés dans l'étude d'impact préalable à la loi pour une République numérique et destinés à alimenter le futur « Service public de la donnée ».

Art L.321-4.-1. II du code des relations entre le public et l'administration

« sont des données de référence les informations publiques mentionnées à l'article L. 321-1 qui satisfont aux conditions suivantes :

« 1° Elles constituent une référence commune pour nommer ou identifier des produits, des services, des territoires ou des personnes ;

« 2° Elles sont réutilisées fréquemment par des personnes publiques ou privées autres que l'administration qui les détient ;

« 3° Leur réutilisation nécessite qu'elles soient mises à disposition avec un niveau élevé de qualité. »

Un décret en Conseil d'État doit également dresser la liste des « jeux de données de référence » et désigner les administrations responsables de leur mise à disposition en fonction de critères de qualité préalablement identifiés.

À l'occasion du sommet du Partenariat pour un gouvernement ouvert (7-9 décembre 2016), l'annonce a été faite d'**une dotation de trois millions d'euros pour le service public de la donnée** porté par la loi République numérique. Issu du deuxième programme d'investissements d'avenir (PIA), ce futur financement de trois millions d'euros doit servir « à encourager le développement d'infrastructures de diffusion à haut niveau de disponibilité, y compris des outils technologiques mutualisables de sécurisation de la diffusion, de gestion industrielle des signalements d'erreur, ainsi que l'interopérabilité et la convergence des données de référence entre elles, par exemple en déployant des solutions de type données liées. »

22. L'ensemble des résultats est disponible ici : <https://www.etalab.gouv.fr/consultation-sp>

Préconisation #2

Faire qualifier les données jurisprudentielles de données de référence : parce que l'ouverture de ces jeux de données est considérée comme prioritaire, et qu'elle est associée à une exigence haute en terme de qualité. Il pourrait être intéressant de vérifier que les données jurisprudentielles remplissent les conditions pour être qualifiées de données de référence, ou à défaut leur conférer un statut équivalent.

Identifier le périmètre de l'open data jurisprudentiel

En consacrant le principe d'un open data jurisprudentiel, et donc de la mise à disposition de l'ensemble des décisions de justice « dans un standard ouvert aisément ré-utilisable et exploitable par un système de traitement informatisé », **le législateur crée une rupture entre les modèles « d'accès au droit » (open access) et d'exploitation des données jurisprudentielles (open data).**

Jusqu'alors, la diffusion des décisions de justice reposait a priori sur un double filtre : **le prononcé public de la décision et une sélection préalable** des décisions par les juridictions en fonction de leur pertinence²³.

S'agissant de l'open data des décisions de jurisprudence, seule la condition du prononcé public de la décision serait désormais maintenue. Autrement dit, sont concernées par l'open data toutes les décisions auxquelles il est possible d'avoir accès auprès du greffe des tribunaux.

La condition de publicité des décisions de justice

Si, pour les décisions de l'ordre administratif, l'article 20 de la loi une République numérique n'a pas expressément articulé la mise à disposition des décisions en open data avec les règles existantes relatives à l'accès et à la publicité des décisions de justice, l'article 21 adopte la solution inverse concernant les décisions de l'ordre judiciaire. Cet article dispose en effet que la mise à disposition des décisions est réalisée « sans préjudice des règles relatives à l'accès et à la publicité des décisions de justice ». Pour l'application de cet article, il faudra donc distinguer les décisions de justice qui sont rendues publiquement – y compris après des débats tenus à huis clos – de celles qui sont rendues de manière non publique. Ainsi, c'est le caractère public de la décision et son libre accès aux tiers qui détermineront la mise à

23. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000000626845>

disposition ou non de la décision en Open data, sous réserve de la protection de la vie privée des personnes.

D'une manière générale, les dispositions de l'article 14 du Pacte des droits civils et politiques et de l'article 6§1 de la Convention européenne des droits de l'Homme imposent le principe de la publicité des débats et des décisions, tout en ouvrant la possibilité aux États d'instituer des exceptions. Dans ce cadre, en droit français, il existe ainsi une série d'exceptions importantes à la publicité des décisions de l'ordre judiciaire qui impliquera que les décisions des matières concernées ne soient pas mises à disposition en Open data.

Il ne peut donc être raisonné globalement. Il convient, pour chaque type de décision, d'analyser les dispositions juridiques applicables en matière d'accès et de publicité afin d'identifier celles qui peuvent figurer parmi les décisions pouvant être mises en Open data. La CNIL a déjà commencé ce travail de recensement. Il est nécessaire de le poursuivre.

Le maintien d'un principe de sélection ?

En matière d'Open data jurisprudentiel, la **suppression de la sélection est au cœur des débats**²⁵.

A l'heure actuelle, elle est exercée de manière discrétionnaire par les juridictions en fonction de règles qui leurs sont propres et **repose sur un principe de pertinence et d'utilité du contenu** de la décision (rareté des faits, première application d'un texte, qualité du raisonnement juridique, etc.). Ainsi, certaines décisions « purement procédurales », ou en séries, ne seront pas publiées car considérées comme sans intérêt pour le lecteur final. On peut également penser à des décisions très standardisées rendues totalement identiques après anonymisation.

Or, si ce filtre se conçoit lorsqu'on raisonne en terme « d'information juridique », le bien-fondé d'une sélection des décisions lorsque l'on se positionne dans le champ de l'Open data est loin d'être évident. En effet, **l'existence même de la décision est, *a minima*, une donnée pertinente en soi**,

.....

25. C'est d'ailleurs le concept même de jurisprudence qui est ici débattu : doit-on l'envisager au sens strict comme un ensemble de décisions distinguées pour leurs qualités particulières (pertinence, rareté des faits, absence de précédent, première application d'un texte, etc.) ou simplement comme une manière de désigner l'ensemble des décisions prononcées par les juridictions administratives et judiciaires ?

peu importe son contenu. La juridiction responsable et la date du prononcé, le sens de la décision (acceptation ou rejet de la demande) sont encore des éléments susceptibles de nourrir des algorithmes prédictifs ou des modèles statistiques.

Au-delà, il faut **prévoir le développement de nouveaux services et usages** grâce à la mise à disposition en Open data des fonds de décisions de justice. Or, il est difficile d'anticiper quelles seront les décisions jugées utiles demain. Au-delà, c'est bien le volume des données qui fait la force de l'Open data.

Les limites encadrant la publication des décisions de justice

En l'état du droit, le caractère « accessible » d'une décision de justice ne préjuge pas forcément de la possibilité de la diffuser²⁶. Bien que cela paraisse peu probable dans la mesure où l'Open data est une « mise à disposition » de données et non une « publication » au sens strict », on peut néanmoins s'interroger sur le point de savoir si ces restrictions pourraient être étendues à l'Open data jurisprudentiel.

Limites imposées par les textes

La diffusion de la jurisprudence est notamment encadrée par la loi sur la liberté de la presse du 29 juillet 1881²⁷. Cette dernière interdit de publier par voie de presse ou « tous autres moyens », des décisions de justice portant sur les procès en diffamation, en cas de remise en cause de la filiation, des actions à fin de subside, des procès en divorce, séparation de corps, les nullités de mariage, les procès en matière d'avortement, d'atteintes à la pudeur, les victimes d'un viol ou bien encore les personnes ayant fait l'objet d'une adoption plénière. La loi exclut également la publicité des poursuites pénales exercées en matière de maladies vénériennes et de nourrice d'enfant. Il en va de même pour les décisions prises à l'égard d'un mineur. Le Code de commerce notamment en son article R123-154 précise que certaines décisions en matière de procédure collective ne sont pas publiables. →

26. <http://www.precisement.org/blog/Se-procurer-la-copie-d-une-decision-de-justice-les-bases-legales-et-un-guide.html>

27. <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006070722>

Limites jurisprudentielles

La jurisprudence a considéré que, pour que la publication d'une décision soit licite, celle-ci ne doit pas être incomplète. Il faut comprendre 1. que les débats judiciaires doivent être reproduits fidèlement et de bonne foi ; 2. qu'il doit être fait mention de l'existence d'un éventuel recours formé dès lors que la décision publiée n'a pas acquis un caractère définitif (pour se prémunir contre l'atteinte à la présomption d'innocence)²⁸.

On le voit, **l'articulation entre les règles de publicité et d'accès, de publication et de sélection des décisions de jurisprudence est une science délicate**. Le basculement vers un open data par défaut va nécessairement interroger la pertinence de l'ensemble de ces règles au vue des caractéristiques spécifiques de ce mode de diffusion (et notamment l'obligation de pallier au risque de ré-identification).

Préconisation #3

Poursuivre le travail de recensement des décisions de justice pouvant être mises à disposition en open data.

28. Voir par exemple : <https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=LEGITEXT000006070722>

Protéger la vie privée du justiciable

Enjeu juridique et sociétal, la vie privée est ici mise en danger par le caractère a priori inconciliable de la mise à disposition en open data de documents et de la protection des données personnelles qu'il contient.

Le principe d'anonymisation des décisions de jurisprudence

« L'anonymisation est autant un principe de pertinence qu'un principe de sécurité des données »

CNIL - Atelier Open Case Law du 29 septembre.

Comme tout document contenant des données à caractère personnel²⁹, le traitement des données jurisprudentielles n'est possible que dans trois cas :

- * en cas de consentement des personnes concernées,
- * en cas d'anonymisation préalable,
- * ou lorsqu'une disposition légale le permet.

En matière de décision de justice, c'est la solution de la pseudonymisation qui a été retenue.

* Définitions

A ce jour, ni l'anonymisation, ni la pseudonymisation ne sont des termes définis par la loi.

*** Anonymisation :** le terme s'entend soit au sens large et englobe alors à la fois les techniques de suppression ou de réduction du caractère identifiant

.....
29. Article 2 de la loi Informatique et Liberté : « Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres ».

des données (anonymisation, pseudonymisation, masquage, offuscation, etc.), ou au sens restreint. Dans ce dernier cas, on considère que l'anonymisation est un processus qui permet l'obtention de données ne pouvant être reliées à des personnes par aucun moyen. Ces données sont alors qualifiées d'anonymes et sortent du champ d'application de la loi Informatique et Libertés. Le caractère irréversible de l'altération subie par les données doit rendre impossible toute ré-identification des personnes par quelque moyen que ce soit³⁰.

En pratique, la CNIL applique la doctrine proposée par le Groupe article 29 (G29, le consortium des autorités de protection européennes) dans son avis du 10 avril 2014³¹. Ainsi, un jeu de données est considéré anonyme si une analyse de risques de ré-identification a été effectuée à la satisfaction des autorités compétentes ou s'il est démontré qu'il n'est pas possible :

- ✱ d'isoler/d'individualiser des informations relatives à un seul individu ;
- ✱ de relier/corréler les données d'un même individu ou groupe d'individus ;
- ✱ de déduire/d'inférer d'un ensemble d'attributs la valeur d'un autre attribut.

Pseudonymisation : la pseudonymisation est un procédé par lequel des données directement identifiantes sont rendues indirectement identifiantes (par exemple en remplaçant un nom par une chaîne de caractères aléatoires). La pseudonymisation est une mesure de sécurité visant à préserver la confidentialité des données. Toutefois, ce processus est potentiellement réversible. On ne peut donc pas qualifier les données obtenues d'anonymes.

La pseudonymisation est la solution qui semble privilégiée par le législateur européen, et celle qui est mise en œuvre sur le territoire français. Elle a pour but d'empêcher la ré-identification directe d'une personne physique par les données intrinsèques à une décision de justice.

Dans sa délibération du 29 novembre 2001 portant recommandation sur

.....

30. Pour les éléments qui suivent, nous avons utilement consulté le support de Féli-cien Vallet « Présentation de la doctrine technique de la CNIL concernant l'anonymisa-tion », daté du 29 novembre 2016.

31. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_fr.pdf

la diffusion de données personnelles sur internet par les banques de données de jurisprudence, la CNIL recommande ainsi d' « anonymiser (...) le nom et l'adresse des parties et des témoins, dans tous les jugements et arrêts librement accessibles sur Internet, quels que soient l'ordre ou le degré de la juridiction et la nature du contentieux, mais cela seulement.³² »

Toutefois, comme indiqué précédemment, la notion d'anonymisation a évolué depuis cette délibération. Les avancées mathématiques, l'augmentation de la puissance de calcul ainsi que le nombre croissant des sources d'informations disponibles sont autant de facteurs ayant contribué à rehausser les risques de ré-identification.

✱ Méthodologie

L'un des impacts majeurs que pourrait avoir la loi Lemaire pour les producteurs de données serait la **systématisation de la pseudonymisation native** des décisions à travers la normalisation des jugements et arrêts. Autrement dit cela imposerait aux magistrats, au moment de la rédaction de leur décision, de prendre en compte l'impératif de protection de la vie privée, et de choisir leur formulation de telle sorte que la l'anonymisation ne vienne pas compromettre l'intelligibilité de la décision.

Si cela constitue probablement la solution la plus simple et la plus fiable en terme de protection des données personnelles, elle est **pour l'instant difficilement envisageable** à court terme d'un point de vue technique, organisationnel, humain et financier, particulièrement pour les juridictions de l'ordre judiciaire et notamment celles du premier degré (du fait de leur nombre et de leur variété).

Le traitement peut en effet être distribué sur toute la chaîne de production.

- ✱ **L'écriture en « privacy by design »** : le risque de ré-identification est pris en compte dès la rédaction de la décision, par exemple grâce un outil d'aide à la rédaction. C'est la solution décrite ci-dessus.
- ✱ **La pré-pseudonymisation** : dans ce scénario, la pseudonymisation n'intervient qu'une fois la décision de justice signée, au niveau des greffes.
- ✱ **La pseudonymisation centralisée** : dans ce scénario, la pseudonymisation repose sur un acteur tiers, qu'il s'agisse de la juridiction suprême ou d'un diffuseur public ou privé.

.....
32. <https://www.legifrance.gouv.fr/affichCnil.do?oldAction=rechExpCnil&id=CNILTEXT000017653503>

L'approche par les risques

Afin de **préserver l'équilibre entre les exigences contraires d'ouverture des données et de protection de la vie privée**, les articles 20 et 21 de la loi pour une République numérique ont **consacré une approche par les risques** : les données jurisprudentielles sont « mises à la disposition du public à titre gratuit dans le respect de la vie privée des personnes concernées. Cette mise à disposition du public est précédée d'une analyse du risque de ré-identification des personnes. »

Cette disposition anticipe également les conséquences de l'article 9 du Règlement européen sur la protection des données personnelles, qui impose cette analyse d'impact à partir de 2018 pour les « [...] données à caractère personnel relatives à des condamnations pénales et à des infractions visées [...] »³³.

La consécration de cette approche équivaut à **renoncer au principe de recherche du « zéro-défaut »** et permet de dépasser ses limites :

- ✱ aujourd'hui, les moyens techniques disponibles rendent **illusoire de garantir la fiabilité absolue ou le caractère irréversible de l'anonymisation** ;
- ✱ au-delà, cette approche permet **un arbitrage au cas par cas** entre les impératifs de respect de la vie privée des personnes citées dans la décision et le maintien de l'intelligibilité de la décision, indispensable à son utilité sociale ou économique : en effet, en matière de jurisprudence, la suppression ou pseudonymisation systématique de l'ensemble des données identifiantes dénature le sens du texte en brouillant les rôles des parties ou en supprimant des faits indispensables à la compréhension de l'espèce.

Le risque doit donc se comprendre comme **l'équilibre entre** :

- ✱ le risque de croisement des données ou d'inférence³⁴,
- ✱ le risque de produire des données insatisfaisantes ou inexploitable.

.....
³³. <https://www.legifrance.gouv.fr/affichCnil.do?&id=CNILTEXT000017653503>

L'évaluation par le risque permettra notamment de différencier les contentieux : par exemple, en matière de propriété industrielle le cas notamment des marques éponymes est à prendre en considération.

✱ Le contenu de l'analyse de risque

Cette analyse s'entend comme **une mise à plat de la chaîne de production des décisions de justice** et devra lister :

- ✱ les sources du risque,
- ✱ les événements redoutés (ici, la ré-identification),
- ✱ la gravité de leur survenance et les typologies de dommages possibles.

Elle contiendra nécessairement **une partie technique décrivant le système de traitement et de pseudonymisation des données** (manuelle, automatisée, assistée). Mais l'analyse devra également porter plus largement sur l'ensemble de la chaîne, du producteur (collecte, statut des agents, information sur les déposants) au réutilisateur (quelle trace laissée par les réutilisateurs, quel contrat, quelle responsabilité ?).

Si l'open data se moque de la finalité du traitement, on peut imaginer faire preuve de « differential privacy » au moment de l'analyse de risque, en prenant en compte les finalités.

En tout état de cause, l'entité qui se voit confier la responsabilité de l'analyse peut, dès lors qu'elle justifie sa décision, décider de **refuser la diffusion de certaines décisions** en raison d'un risque trop élevé, ou **créer des "bulles de confiance"** avec certains acteurs.

Se posera également la question de la périodicité de l'analyse, afin de prendre en compte les évolutions techniques et technologiques.

34. <http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:32016R0679&from=FR>

Préconisation #4

Définir une méthodologie d'analyse d'impact qui soit unique et applicable à l'ensemble des juridictions, et conforme avec l'article 35 du Règlement général sur la Protection des Données personnelles.

✱ L'auteur de l'analyse et la chaîne de responsabilité

L'analyse de risque repose d'abord sur le producteur de la donnée. Mais on doit concevoir que certains intermédiaires qui réclament l'accès à une donnée intègre soient automatiquement tenus comme solidairement responsables du risque de ré-identification.

Au-delà, on pourrait même envisager que cette notion de « chaînage de responsabilité » englobe les « téléchargeurs » des jeux de données afin d'être en mesure de remonter la chaîne des accès (sur un modèle : « une donnée en échange d'un jeu de données »).

L'ensemble du processus doit être itératif, afin de vérifier et lister les décisions dérogatoires, en se fondant sur l'intelligence collective.

Préconisation #5

Déterminer clairement la chaîne de responsabilité en fonction des différents scénarios de réutilisations des données et des usages propres à chaque acteur.

✱ La publicité de l'analyse

Les articles 20 et 21 de la loi pour une République numérique ne prévoient pas la publication de l'analyse de risque de ré-identification. Néanmoins, on pourrait concevoir qu'en cas de non diffusion, les éléments déterminants de l'analyse fassent l'objet d'une communication au public. Il faudra néanmoins prendre en compte la nécessité de préserver un certain « secret technique » en évitant d'exposer des vulnérabilités, ce qui compromettrait la sécurité du dispositif.

Pack de conformité Cnil

La CNIL travaille à la création d'un pack de conformité officiellement annoncé le 29 septembre lors de l'atelier 1 - Pseudonymisation du programme Open Case Law.

Objectif

Ce pack vise à construire collaborativement l'articulation entre la pléthore de textes et de principes, parfois contradictoires, qui s'appliquent en matière d'open data jurisprudentiel. Chaque acteur sera sollicité pour proposer des solutions afin de rechercher un équilibre et de faciliter l'industrialisation de la diffusion des décisions de justice, tout en garantissant la protection des administrés concernés.

Au-delà, le pack de conformité veut redistribuer les responsabilités de chacun dans le cadre du partage d'un « bien commun », et repositionne "l'usage" au cœur des réflexions.

Méthodologie

Un premier état des lieux des besoins et des modes de granularité sera nécessaire. Une lettre de mission auprès du SGG a permis d'officialiser une collaboration entre le SGG, ETALAB et la CNIL sur cette question : il faut d'abord décrire la donnée et son schéma de diffusion.

Impact

On peut imaginer que les acteurs respectant les préconisations de ce pack bénéficient d'une présomption favorable s'agissant de l'analyse de risque de ré-identification.



Autorité de certification

La loi Lemaire renforce et élargit les compétences de la CNIL dans ses missions, au nombre desquelles la certification de la conformité des processus d'anonymisation des données personnelles, dans la perspective de leur mise en ligne et de leur réutilisation. L'anonymisation des bases de données est une condition essentielle à leur ouverture ou à leur partage : elle permet de prémunir les personnes des risques de ré-identification, et les acteurs (administrations émettrices de données, réutilisateurs, entreprises privées qui réalisent des recherches notamment statistiques), de la mise en cause de leur responsabilité en la matière. La certification ou l'homologation de méthodologies d'anonymisation ainsi que la publication de référentiels ou de méthodologies générales par la CNIL sera ainsi un gage de protection des personnes et de sécurité juridique pour les acteurs.

Préconisation #6

***Travailler avec la CNIL pour publier
au plus vite le pack de conformité
sectoriel en matière de données
juridiques et judiciaires et lui sou-
mettre les méthodologies d'anony-
misation pour homologation.***

Réflexions sur les données « pseudo-nymisables » (sous-atelier du 29/9)

En considérant qu'on dispose demain des décisions de justice sous une forme numérique, on peut détailler plus précisément dans un fichier **les éléments candidats à la pseudonymisation** : les données d'une part, autrement dit le contenu-même de la décision qui avec un formalisme variable selon les juridictions inclut des en-têtes avant le texte de la décision, et les méta-données d'autre part.

Les métadonnées peuvent par exemple couvrir les indications suivantes, et reprendre tout ou partie des éléments présents dans l'en-tête :

- * Juridiction
- * Nature des contentieux (nomenclature)
- * Degrés de juridiction
- * Lieu de la juridiction
- * Descripteurs (plan de classement jurisprudentiel)
- * Identifiant ECLI
- * Chambre
- * Nom des parties
- * Date de prononcé de la décision

En l'état du droit antérieur à la loi pour une République numérique, seules les métadonnées concernant le nom des parties doivent être anonymisées au moment de la diffusion, à l'inverse des noms des magistrats ou des avocats des parties qui ne sont pas concernés par l'obligation d'anonymisation (sauf s'ils sont partie à la décision).

L'atelier sur l'anonymisation (cf partie préliminaire) a été l'occasion de **recenser l'usage sur les données identifiantes figurant classiquement dans une décision de justice** et susceptibles de donner lieu à croisement (pratique de pseudonymisation OUI/NON) :

	Ordre Administratif	Ordre Judiciaire	DILA / Légifrance	Défenseur des Droits	INPI
<i>Données contenues dans la décision</i>					
Noms des parties	OUI	OUI	OUI	OUI	OUI
Demandeur	OUI	OUI	OUI	OUI	OUI
Défendeur	OUI	OUI	OUI	OUI	OUI
Fonction de la personne	NON	NON	NON	OUI	NON
Personne morale (en cas d'évidence de la personne physique)	OUI		NON	OUI	NON
Personne morale	NON	NON	NON	OUI	NON
Personne morale unipersonnelle	NON (sauf si Nom PP)				
Parties à l'instance (PP)	OUI	OUI	OUI		
Magistrats	NON	NON	NON		
Avocats	NON	NON	NON		
Avocats/Mandataires (partie à l'instance)	OUI	OUI	OUI	NON	OUI
Auxiliaires de justice	NON	NON	NON		
Experts / cf. secret médical / retrouver le lieu	NON	NON	NON	OUI	
Prénom	OUI		NON	OUI	OUI
Surnom					
Nom et Prénom pour les mineurs			OUI		
Civilité	NON				
Lieu de la juridiction					
Lieu des faits					
Date de naissance					
Adresse physique (Personne Physique)	OUI	OUI	OUI	OUI	OUI
Adresse électronique (Personne Physique)	OUI	OUI	OUI	OUI	OUI
Adresse (Personne Morale)	NON		NON		

Garantir la qualité des données et leur facilité de réutilisation

Les critères de qualité de l'open data

Le principe d'open data par défaut augmente considérablement la **volume des données concernées**, ce qui constitue un défi en soi pour les producteurs puisqu'il s'agira de garantir une qualité optimale égale pour un périmètre étendu des données de jurisprudence.

L'adoption des **standards et bonnes pratiques de l'open data** (identifiant ECLI et format numérique ouvert et interopérable) implique que les chaînes de traitement adéquates soient mises en place, et que **les règles de génération d'un ECLI** soient posées.

Critères de qualité de l'open data

Les données mises à disposition en open data doivent répondre à des critères de qualité. Un consensus dans la communauté des développeurs s'attache aux aspects suivants (entre huit et dix critères généralement, cette liste n'étant pas limitative).

Un jeu de données en open data doit être exhaustif (1), et le dépôt permanent et stable (2).

Les données doivent être le plus fidèles possibles à la donnée brute, non agrégées par exemple (3).

Les données doivent être à jour ; on peut parler de fraîcheur des données versées en open data (4).

Elles sont rendues accessibles sous des formats le plus universels possibles, d'un point de vue technique (5). Ces formats de fichiers ne sont pas des formats propriétaires (6).

Les données sont lisibles et exploitables par une machine (7).

La mise à disposition ne peut être discriminatoire, c'est-à-dire que

→

l'on ne saurait en réserver l'accès à des catégories de publics et pas d'autres (8).

Elles sont distribuées sous licence ouverte (9) et sont gratuites (10).

On peut aussi ajouter que les données et notamment leur modèle doivent être dûment documentées (par exemple, si mise à disposition en XML : une DTD, ou modèle de données, doit être fournie). D'autres considérations techniques encore peuvent être convoquées.

Un consensus s'est établi parmi les participants au programme autour des critères de qualité les plus décisifs pour la donnée jurisprudentielle : la perspective de données de référence garanties par l'acteur public est perçue par tous comme un gain, notamment pour garantir une anonymisation homogène et de qualité, la fraîcheur et la périodicité des livraisons sont la première préoccupation, une transmission en temps réel est même évoquée.

Enfin, les attentes portent sur l'exhaustivité et le périmètre prévisionnel des données concernées par la loi République numérique : le basculement en open data des décisions de première instance est très attendu, notamment les jugements des tribunaux de commerce, des conseils de prud'hommes, et bien entendu des TGI. En complément, la communauté gagnerait à ce que les divers référentiels de l'organisation administrative soient ouverts en open data : par exemple, la liste des juridictions avec leur siège et leur ressort.

Les métadonnées de jurisprudence : faciliter la réutilisation des décisions

Les données de jurisprudence sont une ressource textuelle qui n'est pas réutilisable à leur émission en l'état (copie de l'arrêt du greffe). C'est là qu'interviennent les métadonnées³⁵. Aujourd'hui, ces dernières sont l'objet d'une volonté de normalisation au niveau européen, à travers le European Case Law Identifier (ECLI), norme communautaire de description des métadon-

35. Les métadonnées sont « des données sur les données », qu'elles décrivent et qualifient.

nées des décisions de justice. Son cadre fonctionnel a été fixé en 2011 par le Conseil de l'Union européenne³⁶.

Contenu et métadonnées prévues par la norme ECLI

« L'identifiant européen de la jurisprudence (European Case Law Identifier – ECLI) a été conçu pour contribuer à la citation correcte et sans équivoque des décisions de justice rendues par les juridictions européennes et nationales. Un ensemble de métadonnées uniformes améliorera les outils de recherche de la jurisprudence. »³⁷

Autrement dit, l'identifiant ECLI sert à identifier une décision de façon univoque, non seulement pour des besoins de citation rigoureuse mais aussi pour en permettre la lecture par une machine ; ECLI, comme ELI (European Legislation Identifier), sont des concepts issus du web sémantique et sont pour partie un jeu de métadonnées descriptives, mais aussi un URI (unique resource identifier) qui permet de désigner de façon non ambiguë une ressource textuelle, et d'y accéder.

L'identifiant européen de la jurisprudence comporte les éléments suivants, séparés entre eux par deux-points :

- * l'abréviation ECLI, en guise de préfixe ;
- * le code du pays (FR pour la France) ;
- * le code de la juridiction qui a rendu la décision ;
- * l'année de la décision³⁸ ;
- * un numéro d'ordre, de 25 caractères alphanumériques maximum,
- * présenté selon un format décidé par chaque État membre. Le
- * numéro d'ordre peut contenir des points («.») mais aucun autre
- * signe de ponctuation.

.....
36. Conclusions du Conseil préconisant l'introduction d'un identifiant européen de la jurisprudence et un ensemble minimal de métadonnées uniformes pour la jurisprudence, [http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52011XG0429\(01\)](http://eur-lex.europa.eu/legal-content/FR/ALL/?uri=CELEX:52011XG0429(01)), Office des Publications de l'Union européenne, consulté le 5 novembre 2016

37. [http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52011XG0429\(01\)&from=FR](http://eur-lex.europa.eu/legal-content/FR/TXT/HTML/?uri=CELEX:52011XG0429(01)&from=FR)

38. Cf page suivante

La France fait partie des pays ayant adopté l'ECLI³⁹, et le Secrétariat général du Gouvernement a désigné la DILA comme coordinateur national ECLI.

Les cours suprêmes (Conseil constitutionnel, Cour de cassation, Conseil d'État) identifient déjà leurs décisions avec un ECLI. Conseil constitutionnel et Conseil d'État génèrent eux-mêmes l'identifiant ECLI de leurs décisions, avant de les transmettre par fichier à la DILA, qui les enregistre dans ses bases CONSTIT et JADE respectivement. Quant à la Cour de cassation, elle identifie ses arrêts par un identifiant interne, le numéro d'affaire, puis fournit les éléments de base (mais pas le numéro de pourvoi) à la DILA, qui construit ensuite l'ECLI⁴⁰.

Ces décisions sont affichées, après anonymisation, avec leur ECLI, aussi bien sur le portail propre de chaque juridiction que sur le portail Legifrance.

La DILA complète les données obligatoires associées aux ECLI. Ces décisions ainsi enrichies sont affichées, après anonymisation, avec leur ECLI, sur le portail de chaque juridiction. La DILA les diffuse sur Legifrance et les met à disposition en Open data via le site data.gouv. La DILA les fournit également à la Direction Justice de la Commission européenne sous un format XML adhoc pour mise à disposition sur le portail européen.

.....

38. Le code juridiction doit, entre autres : 1) comporter au moins un et au plus sept caractères 2) toujours commencer par une lettre mais peut aussi contenir des chiffres 3) être choisi de manière à sembler logique aux personnes familiarisées avec l'organisation de l'appareil judiciaire du pays concerné 4) au minimum correspondre à une abréviation du nom de la juridiction, tout en pouvant aussi contenir une indication de la chambre ou de la section concernée au sein de cette juridiction, notamment s'il est habituel de désigner la chambre ou la section dans les renvois à la jurisprudence du pays concerné 5) ne devrait pas contenir d'information relative au type de document.

39. Chaque État membre reste en effet libre d'adopter ou non la norme ECLI.

40. La DILA a pu reconstituer des ECLI rétroactivement jusqu'en 2008, mais pas avant cette date, car le numéro d'affaire n'était pas fourni dans le fichier remis jusqu'alors.

Préconisation #7

La France ne fait actuellement pas partie du groupe d'États-Membres constitué autour du projet « Building on ECLI » lancé en octobre 2015. Or, l'importance de l'enjeu entourant la circulation et l'interopérabilité des décisions de justice, mais aussi la possibilité de bénéficier du partage de bonnes pratiques et du transfert de compétences entre États-Membres militent en faveur d'une participation française aux travaux du groupe BO-ECLI.

Préconisation #8

Dresser l'inventaire détaillé des métadonnées utiles et nécessaires pour les décisions du fonds, selon une matrice à définir (type juridiction et/ou par type de contentieux) et normaliser ces métadonnées. Ces décisions devraient a minima respecter la norme ECLI.

✱ Proposition de référentiels pour un identifiant ECLI des décisions du fonds

Dans le cadre du deuxième atelier du programme Open Case Law, la DILA a proposé un recensement des codes juridictions à ajouter dans la syntaxe de l'identifiant ECLI pour prendre en compte l'ouverture des décisions des juridictions du fonds.

Ce recensement ne traite donc pas des cours suprêmes. En revanche, sont traitées les juridictions suivantes : les tribunaux de commerce, les cours d'appel, les tribunaux d'instance, les tribunaux de grande instance, les cours administratives d'appel, les tribunaux administratifs, les conseils de prud'hommes. Il a été choisi 4 caractères pour la ville où siège la juridiction et 2 ou 3 caractères pour le type de juridiction. Sur les 908 juridictions recensées, 902 ont un code valide et seules six ont un code en « conflit », c'est-à-dire que leurs codes sont communs à deux juridictions (aperçu ci-dessous).

Roche-sur-Yon la	LARY	LARYCPH	code ok
Roche-sur-Yon la	LARY	LARYTC	code ok
Roche-sur-Yon la	LARY	LARYTGI	code ok
Roche-sur-Yon la	LARY	LARYTI	code ok
Rodez	RODE	RODETC	code ok
Rodez	RODE	RODETGI	code ok
Rodez	RODE	RODETI	code ok
Rodez	RODE	RODECPH	code ok
Romans	ROMA	ROMATC	code ok
Romans	ROMA	ROMATI	code ok
Roubaix	ROUB	ROUBTI	code ok
Roubaix	ROUB	ROUBCPH	code ok
Rouen	ROUE	ROUECA	code ok
Rouen	ROUE	ROUETC	code ok
Rouen	ROUE	ROUETGI	code ok
Rouen	ROUE	ROUETI	code ok
Rouen	ROUE	ROUECPH	code ok
Rouen	ROUE	ROUETA	code ok

Extrait du référentiel proposé par la DILA (disponible sur demande)

Bouleversement des usages et des business models

« 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 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. »⁴¹

La mise à disposition gratuite et sous une forme réutilisable bouleverse le modèle économique actuel de diffusion des données jurisprudentielles, présageant d'un manque à gagner chez certains acteurs actuels diffuseurs de ces données comme les éditeurs juridiques, mais permettant aussi **l'émergence de nouveaux services attachés à cette donnée,** à l'initiative de nouveaux acteurs tels que les legal start-ups.

L'avènement de l'open data pour les données publiques est un défi pour les éditeurs juridiques, particulièrement quand les modèles économiques ont reposé de longue date sur la difficulté pour les professionnels du droit, premiers clients de ces éditeurs, d'accéder à un volume significatif de données de jurisprudence. Cela reste d'actualité par exemple pour les décisions des cours d'appel, dont la sélection disponible en open access sur Legifrance et en open data sur data.gouv.fr est très réduite par rapport au flux réel de ces décisions (de l'ordre de 4% de la production totale⁴²).

.....
41. Jean-Paul Jean, Président de chambre à la Cour de cassation, directeur du Service documentation, des études et du rapport, lors du colloque du 14 octobre consacré à "La jurisprudence dans le mouvement open data" (transcription accessible ici : <https://www.courdecassation.fr/IMG/Jean-Paul%20Jean%20%20Open%20data.PDF>)

42. Mise au point sur l'open data des décisions de justice au 22 juin 2016, par Michaël Benesty, post publié sur medium.

Ainsi, les PDF des arrêts transmis par les cours d'appel sont numérisés, nettoyés, convertis dans un format XML selon un modèle de données propriétaire, et le cas échéant enrichis de métadonnées descriptives voire d'autres éléments propres à chaque éditeur, par exemple des liens vers sa doctrine, ou encore une indexation thématique qui correspond à son plan de classement documentaire.

Les éditeurs s'inscrivent alors naturellement dans la chaîne de diffusion de la donnée jurisprudentielle, prenant à leur charge la transformation de la donnée, son enrichissement et enfin sa publication et apparaissent comme des tiers indispensables en relais de l'acteur public, producteur de ces données.

Ce rapport se trouve désormais chahuté par la perspective d'une mise à disposition massive des données de jurisprudence et pour toute la communauté.

Dans le même temps, c'est aussi **une opportunité nouvelle** pour qui veut explorer ces données au moyen de technologies empruntées aux architectures big-data et d'algorithmes de machine learning ou d'analyse statistique.

On observe aujourd'hui **une convergence inédite entre le mouvement d'ouverture des données publiques et la maturité de ces technologies** dites de big data combinées à l'intelligence artificielle.

L'évolution législative vient ici opportunément servir le marché émergent de la Legal Tech, qui a vu fleurir en 2016 une variété d'acteurs et d'initiatives pour certaines très disruptives et prometteuses quant à l'amélioration de l'accès à la donnée juridique. L'open data et ses principes d'égalité dans l'accès aux données devraient permettre que ces nouveaux acteurs trouvent désormais leur place dans l'écosystème du droit.

Conclusion

Les différentes questions identifiées dans ce livre blanc seront traitées dans les textes d'application de la loi pour une République numérique, actuellement en cours de rédaction au sein de la Chancellerie⁴³. Ils viendront préciser le cadre de l'open data jurisprudentiel « entre service public et nouveau marché »⁴⁴.

A travers les échanges avec les multiples acteurs impliqués dans le circuit de production, de diffusion et de consommation des données jurisprudentielles. l'atelier 3 sur le service public de la donnée juridique a été l'occasion de « diagnostiquer » les attentes et les craintes associées à l'open data en matière de décisions de justice.

Le tableau de synthèse suivant est un premier résumé de ces enjeux et des principaux points de vigilance par catégorie d'acteurs, situés dans leurs fonctions au sein du circuit de diffusion des décisions de justice.

.....

43. Au moment de la rédaction de ce livre blanc, les services du ministère de la Justice n'ont pas encore arbitré sur le point de savoir si les dispositions relatives aux deux ordres de juridiction allaient être réunies au sein d'un même décret ou faire l'objet de deux textes distincts.

44. Extrait du discours de M. Jean-Paul Jean, Président de chambre à la Cour de cassation, directeur du Service de documentation, des études et du rapport <https://www.courdecassation.fr/IMG/Jean-Paul%20Jean%20%20Open%20data.PDF>

Préconisation #9

Proposer aux services du ministère de s'inspirer de la méthode de rédaction collaborative des textes réglementaires développée dans le cadre d'Open Law⁴⁵. A minima, ne pas rompre le dialogue engagée avec les acteurs de l'Open data judiciaire dans le cadre de ce programme et les associer au différentes étapes de la rédaction.

45. <https://hackpad.com/Mthodologie-de-co-rdaction-des-textes-rglementaires-relatifs-la-dmatrialisation-des-procdures-4.1-MzjwFHWcFFd>

Synthèse générale

En tant qu'acteurs	Je produis la donnée		Je délivre la donnée			J'utilise la donnée		Quels impacts de la loi Lemaire? Enjeux, attentes et points de vigilance
	Emission	Enrichissement	en Open Data	en Open Access	en accès payant	Consultation	Exploitation	
Cours Suprêmes	OUI	OUI	NON	OUI	OUI	OUI	NON	Enjeux : Gestion de la volumétrie, Données personnelles Attentes : - Nouvelles exploitations de la donnée jurisprudentielle, notamment à des fins statistiques - Accès à un grand nombre de décisions de justice qui se rapprochent au plus près des faits d'espèce (juridictions du fond) - Nouveaux services innovants d'accompagnement à la prise de décision (notamment modélisation du risque, montant des indemnités, etc.) Points de vigilance: - Que le big data et la justice prédictive appauvrissent le rendu de la justice à travers une automatisation non maîtrisée - Que le big data conduise à la notation et la modélisation du comportement des magistrats - Un formalisme d'écriture lourd à mettre en place (anonymisation native) et une gestion des flux difficiles
Les Juridictions du fonds	OUI	OUI	NON	NON	NON			
Professionnels du Droit (hors magistrats)	NON	NON	NON	NON	NON	OUI	NON*	Enjeux : exhaustivité et intégrité des données disponibles, nouveaux usages Craintes : - Perdre en qualité de données sur les accès gratuits (ex : si l'open data réduit le nombre de décisions sur légitimité) - Que le big data conduise à la notation et à la mesure de la performance des avocats + risque de ré-identification - Devoir financer au prix fort l'accès aux données
DILA	NON	OUI	OUI	OUI	NON**	NON	NON	Enjeu : - Volumétrie et qualité des données - Adoption des standards de l'Open Data et des recommandations européennes (H2020, dont ECLI)

En tant qu'acteurs	Je produis la donnée		Je délivre la donnée				J'utilise la donnée		Quels impacts de la loi Lemaire? Enjeux, attentes et points de vigilance
	Emission	Enrichissement	en Open Data	en Open Access	en accès payant	Consultation	Exploitation		
Etab	NON	NON	OUI	NON	NON	NON	NON	Enjeux : - Volumétrie et qualité des données - Adoption des standards de l'open data - Emergence de nouveaux usages	
Editeurs	NON	OUI	NON	NON	OUI	OUI	OUI	Enjeux : - Emergence de nouveaux business models et nouveaux usages - Volumétrie et qualité des données en open data (inclus la qualité de l'anonymisation) Points de vigilance : - Données personnelles (sur le stock avant ouverture en open data) Attentes : - Conserver un accès aux données intégrés - Qualité des services attachés à la donnée mise à disposition (API ou web service pour l'open data) - Exhaustivité, fraîcheur et intégrité des données disponibles	
Acteurs Legaltech***	NON	OUI	NON	OUI	OUI	NON	OUI	Enjeux : - Emergence de nouveaux business models et nouveaux usages - Volumétrie et qualité des données en open data (inclut la qualité de l'anonymisation) Attentes : - Qualité des services attachés à la donnée mise à disposition (API ou web service pour l'open data) - Exhaustivité, fraîcheur et intégrité des données disponibles Craintes : - Une régulation trop restrictive autour des services innovants (prédictibilité)	

En tant qu'acteurs	Je produis la donnée		Je délivre la donnée			J'utilise la donnée		Quels impacts de la loi Lemaire? Enjeux, attentes et points de vigilance
	Emission	Enrichissement	en Open Data	en Open Access	en accès payant	Consultation	Exploitation	
Universitaires (étudiants, enseignants, chercheurs)	NON	NON	NON	NON	NON	OUI	OUI****	Enjeux : • Volumétrie et qualité des données en open data • Qualité des services attachés à la donnée mise à disposition (API ou web service pour l'open data, moteur de recherche pour l'open access) • Exhaustivité et intégrité des données disponibles Attentes : • Maintien de la gratuité des licences à des fins de recherche • Accès à des jeux de données plus larges et représentatifs donc scientifiquement plus fiables (notamment décisions de justice de première instance) • Facilitation des travaux de recherche en Droit <i>NB : les universitaires ont accès aux bases de données des éditeurs juridiques via l'abonnement payant des bibliothèques universitaires, et à celles des diffuseurs publics sur la base de licence gratuites</i>
Justiciables	NON	NON	NON	NON	NON	OUI	NON	Attentes : • Garanties accrues sur la protection des données personnelles (respect de la vie privée), notamment par une chaîne de responsabilité et des voies de recours claires en cas de manquement • Meilleur accès au droit • Émergence de nouveaux usages Craintes : • Dérive possible vers un mode de justice automatisé ou régi par une approche par le risque

En tant qu'acteurs	Je produis la donnée		Je délivre la donnée			J'utilise la donnée		Quels impacts de la loi Lemaire? Enjeux, attentes et points de vigilance
	Émission	Enrichissement	en Open Data	en Open Access	en accès payant	Consultation	Exploitation	
Société civile et tiers secteurs	NON	NON	NON	NON	NON	OUI	OUI	Enjeux : Adoption des standards de l'open data, Exhaustivité et fraîcheur des données Attentes : - Garanties accrues sur la protection des données personnelles (respect de la vie privée) - Facilitation des démarches de "vigilance citoyenne", garantie d'une meilleure transparence
Inpi	OUI	OUI	OUI	OUI	NON	OUI	OUI	Enjeux : - S'adapter aux nouveaux standards des données (ECL...) - Faire évoluer techniquement la diffusion (API) Points de vigilance : - Le service rendu ne devra pas être dégradé par rapport au service actuel (anonymisation réductrice, accès aux données retardées...) Attentes : - Les décisions des TGI devraient pouvoir être diffusées au format numérique

*Sauf cabinet en capacité d'exploiter les base de données pour développer des services innovants en interne

** La diffusion est gratuite, mais elle s'accompagnait, pour certains réutilisateurs, de conventions de services payantes qui ont pris fin au 31 décembre 2016

*** notamment sociétés prestataires proposant des outils d'anonymisation (ex : Témis) ou un enrichissement des décisions de justice (Ex : doctrine.fr)

**** par exemple pour des recherches en sociologie du droit ou en histoire du droit, des études statistiques, etc.

NB: les acteurs Legal Tech incluent au global les start-ups mais aussi les éditeurs et les grands cabinets qui ont une gestion des données outillée (knowledge management) voire créent des services et solutions logicielles autour des données publiques.

Coordination éditoriale : Camille Charles

Conception graphique et maquette : Mathilde Roussillat Sicsic

Cet ouvrage est disponible en version
numérique sur le site www.openlaw.fr

Livre blanc diffusé sous licence Cc by SA 4.0



