



HAL
open science

La réutilisation des données de la recherche après la loi pour une République numérique

Lionel Maurel

► To cite this version:

Lionel Maurel. La réutilisation des données de la recherche après la loi pour une République numérique. La diffusion numérique des données en SHS - Guide de bonnes pratiques éthiques et juridiques, Presses Universitaires de Provence, 2018, 9791032001790. hal-01908766

HAL Id: hal-01908766

<https://hal.science/hal-01908766v1>

Submitted on 30 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

La réutilisation des données de la recherche après la loi pour une République numérique

Par Lionel Maurel. Décembre 2017.

La question des données de la recherche devient de plus en plus prégnante aujourd'hui, à mesure notamment que le concept d'*Open Science* (science ouverte) s'établit comme un nouveau paradigme pour l'activité scientifique. Au-delà des publications issues du processus de recherche, ce sont désormais les données sous-jacentes qui retiennent une attention croissante. L'enjeu de la réutilisation de ces données de la recherche devient alors crucial et impose aux acteurs de maîtriser le cadre juridique applicable.

Celui-ci a connu depuis la fin de l'année 2015 une refonte en profondeur avec l'adoption en France de deux lois : la loi relative à la gratuité et aux modalités de la réutilisation des informations du secteur public¹ (dite « loi Valter ») et la loi pour une République numérique² (dite « Loi Lemaire »). Les dispositions combinées de ces deux textes ont conduit à la mise en place d'un principe d'ouverture ou d'*Open Data* « par défaut », qui modifie fortement les règles gouvernant la réutilisation des informations publiques en France³.

Si ces textes ne traitent pas directement des données de la recherche (ou seulement de manière fugace), ils posent des principes suffisamment généraux pour être applicables aux établissements d'enseignement et de recherche, y compris en ce qui concerne les données produites dans le cadre des activités de recherche proprement dites (I). Ces dernières vont bien dorénavant être intégrées dans le principe d'ouverture par défaut, qui impose une mise en ligne spontanée et une libre réutilisation (II). Ce n'est que dans une série de cas exceptionnels que les données de recherche – comme toutes les données publiques – échapperont à cette règle générale, notamment pour protéger des droits appartenant à des tiers (propriété intellectuelle, vie privée, confidentialité et secrets) (III).

I) L'inclusion des données de recherche dans le champ du principe d'ouverture « par défaut » des données publiques

¹ Loi n° 2015-1779 du 28 décembre 2015 relative à la gratuité et aux modalités de la réutilisation des informations du secteur public :

<https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000031701525&categorieLien=id>

² Loi n° 2016-1321 du 7 octobre 2016 pour une République numérique :

https://www.legifrance.gouv.fr/affichTexte.do?jsessionid=11B2870334A39A7F45DC7C4AC3E3F8FF.tplgfr35s_1?cidTexte=JORFTEXT000033202746&categorieLien=id

³ Cette inclusion des données de la recherche dans le principe général d'ouverture par défaut a été confirmée par un rapport publié dans le cadre de la BSN. Voir : Ouverture des données de recherche - Guide d'analyse du cadre juridique en France V2, 11 avril 2018 [En ligne]

: <http://www.bibliothequescientifiquenumerique.fr/guide-analyse-cadre-juridique-ouverture-donnees-recherche-v2/>

a) L'instauration d'un principe d'*Open Data* par défaut

Avant 2016, la réutilisation des données publiques était régie en France par la loi du 17 juillet 1978 relative aux relations entre l'administration et les usagers (dite « loi CADA »). La partie de ce texte relative aux « informations publiques » résultait de la transposition dans le droit national de la directive PSI (Public Sector Information) du 17 novembre 2003⁴. Celle-ci instaurait bien un droit à la « libre réutilisation » des informations produites par les administrations, mais sans leur imposer une mise en ligne spontanée des données. Par ailleurs, l'administration restait libre en principe de faire payer les réutilisations par le biais de redevances, avec un encadrement de l'établissement des tarifs.

Le législateur français avait par ailleurs choisi la possibilité – ouverte par la directive de 2003 – de fixer un régime dérogatoire pour deux catégories d'établissements : les établissements, organismes ou services culturels (c'est-à-dire les bibliothèques, archives et musées), ainsi que les établissements et institutions d'enseignement et de recherche. Ces institutions restaient libres de « fixer les conditions dans lesquelles les informations peuvent être réutilisées », en dérogation au principe de libre réutilisation.

Les lois Valter et Lemaire ont profondément modifié le cadre de la réutilisation des données publiques, en instaurant un principe d'ouverture « par défaut » (dit aussi d'*Open Data*) (Chignard et Simon 2012). Adoptée en décembre 2015, la loi Valter établit tout d'abord un principe de gratuité dans la réutilisation des informations publiques. Les administrations perdent la faculté d'établir des redevances pour faire payer les réutilisations (sauf dans des cas exceptionnels⁵) et le texte rend caduques les redevances préexistantes.

De son côté, la loi Lemaire met en effet à la charge des administrations comptant plus de 2 500 agents une obligation de mise en ligne spontanée et dans un format ouvert des documents administratifs qu'elles produisent dans le cadre de leurs activités. Une fois la mise en ligne effectuée, les informations publiques contenues dans ces documents deviennent en principe librement réutilisables par tout un chacun, y compris à des fins commerciales. Le texte énumère à son article 6 les catégories de documents concernés par cette obligation de diffusion spontanée :

[...] lorsque ces documents sont disponibles sous forme électronique, les administrations [...] publient en ligne les documents administratifs suivants :
« 1° Les documents qu'elles communiquent en application des procédures prévues au

⁴ Voir Commission européenne. European legislation on re-use of public sector information :

<https://ec.europa.eu/digital-single-market/en/european-legislation-reuse-public-sector-information>

⁵ Les seules administrations qui peuvent continuer à lever des redevances doivent dégager au moins 25% de ressources propres sur la revente de données, ce qui correspond à des hypothèses rares comme l'IGN ou Météo France. Un décret fixe la liste de ces administrations et il ne contient aucun établissement d'enseignement supérieur ou de recherche. Voir Décret n° 2016-1617 du 29 novembre 2016 relatif aux catégories d'informations publiques de l'État et de ses établissements publics administratifs susceptibles d'être soumises au paiement d'une redevance de réutilisation :

<https://www.legifrance.gouv.fr/eli/decret/2016/11/29/PRMJ1630605D/jo/texte>

présent titre, ainsi que leurs versions mises à jour⁶ ;
« 2° Les documents qui figurent dans le répertoire mentionné au premier alinéa de l'article L. 322-6 ;
« 3° Les bases de données, mises à jour de façon régulière, qu'elles produisent ou qu'elles reçoivent et qui ne font pas l'objet d'une diffusion publique par ailleurs ;
« 4° Les données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental.

On peut déduire de ces dispositions que tout document administratif communicable au sens de la loi CADA a désormais vocation à entrer dans le champ de l'ouverture par défaut. Cette dernière expression signifie que toutes informations publiques ont vocation en principe à être diffusée en *Open Data* et que ce n'est que dans certaines situations exceptionnelles que les administrations seront déliées de cette obligation.

b) La suppression de la dérogation bénéficiant aux établissements d'enseignement supérieur et de recherche

Dès avant l'année 2015, les gouvernements successifs avaient entamé une politique d'*Open Data* au niveau des administrations centrales. Un décret et une circulaire du 26 mai 2011 ont ainsi imposé aux ministères la diffusion de leurs données sur le portail data.gouv.fr en les plaçant sous la licence ouverte/*Open Licence*⁷. Le ministère de l'Enseignement supérieur a rejoint ce mouvement et l'a même prolongé en 2014 par l'ouverture de sa propre plateforme d'*Open Data*, comportant plus de 80 jeux de données en 2017⁸. Néanmoins, cet exemple n'a pas été réellement suivi au niveau des établissements placés sous la tutelle du MESR. Alors que de nombreuses collectivités locales avaient commencé avant la loi Lemaire à développer sur une base volontaire des politiques d'*Open Data*, cela n'a pas été le cas pour les établissements d'enseignement supérieur et de recherche (notamment les universités), hormis quelques démarches isolées.

La loi Lemaire modifie en profondeur la situation en faisant passer ces établissements d'un régime d'ouverture choisie à une ouverture obligatoire. L'ancien article 11 de la loi, qui réservait un statut dérogatoire aux données culturelles et de recherche, a en effet été abrogé par le législateur. Si la loi Valter maintient plusieurs spécificités en ce qui concerne la réutilisation des données des institutions culturelles, ce n'est pas le cas pour les données des établissements d'enseignement supérieur et de recherche.

Cette évolution va concerner l'ensemble des informations publiques que ces établissements produisent (données administratives et financières, données liées aux formations dispensées, données liées à leurs bâtiments, etc.). Mais les données de la recherche, c'est-à-dire celles produites dans le cadre d'activités de recherche proprement dites, seront aussi englobées dans l'ouverture par défaut.

⁶ Ce 1° instaure un *Open Data* « à la demande » : si une administration ne satisfait pas d'elle-même aux obligations d'ouverture par défaut, il suffit de lui adresser une demande de communication d'un document pour qu'elle soit obligée ensuite de le publier en ligne et de rendre librement réutilisables les informations qu'il contient (à condition que le document soit bien communicable).

⁷ Voir *infra*.

⁸ Voir <https://data.enseignementsup-recherche.gouv.fr/pages/home/>

c) Les raisons de l'inclusion des données de recherche dans le principe général

L'inclusion des données de recherche dans ce nouveau principe général d'ouverture peut de prime abord surprendre. Les chercheurs disposent en effet d'un statut particulier visant à leur permettre de contrôler la manière dont leurs productions intellectuelles sont diffusées. Il s'agit notamment de la seule catégorie d'agents publics à conserver pleinement leur droit d'auteur sur les créations qu'ils produisent dans le cadre de leur mission de service public⁹ (articles, ouvrages, cours, supports pédagogiques, etc.). Les prérogatives du droit d'auteur (notamment le droit moral) permettent aux chercheurs de choisir s'ils souhaitent ou non publier leurs écrits, ainsi que la forme que cette publication doit prendre. Lorsqu'ils choisissent la diffusion en ligne de ces créations, ils restent par ailleurs maîtres d'en autoriser ou non la réutilisation (en les plaçant sous licences libres).

Mais il existe une différence importante de régime juridique applicable aux œuvres de type écrits scientifiques et aux données. Si l'on entend le terme « données de recherche » au sens strict (c'est-à-dire d'informations générées dans le cadre d'un processus de recherche), alors le chercheur ne sera pas considéré comme un « auteur », car les données ne sont généralement pas en tant que telles des « œuvres » protégeables par le droit d'auteur. Un droit de propriété intellectuelle spécifique existe néanmoins pour la protection des bases de données¹⁰, mais son fonctionnement est différent de celui du droit d'auteur. Ce droit dit de « producteur de base de données » n'appartient pas à l'origine aux personnes physiques qui réalisent la base, mais à l'entité qui a effectué des « investissements substantiels » pour rendre cette opération possible. Dans la plupart des hypothèses, ce seront donc les établissements de tutelle des chercheurs qui auront la qualité de « producteurs » et posséderont les droits attachés aux bases de données de recherche.

Or la loi République numérique a explicitement « neutralisé » le droit des bases de données des administrations pour faire primer le principe de libre réutilisation. Le nouvel article 11 du texte indique ainsi :

Sous réserve de droits de propriété intellectuelle détenus par des tiers, les droits des administrations mentionnées au premier alinéa de l'article L. 300-2 du présent code, au titre des articles L. 342-1 et L. 342-2 du code de la propriété intellectuelle [c'est-à-dire le droit de producteur de bases de données], ne peuvent faire obstacle à la réutilisation du contenu des bases de données que ces administrations publient en application du 3° de l'article L. 312-1-1 du présent code.

Il en résulte que les données produites par les chercheurs sont bien comprises dans le principe d'ouverture par défaut. La situation sera donc très différente de celles des écrits scientifiques et autres créations produites par les chercheurs dans le cadre de leurs activités.

⁹ Voir Jurispedia. Droit d'auteur des agents publics :

[http://fr.jurispedia.org/index.php/Droits_d'auteur_des_agents_publics_\(fr\)](http://fr.jurispedia.org/index.php/Droits_d'auteur_des_agents_publics_(fr))

¹⁰ Voir APP. Quels sont les droits reconnus au producteur d'une base de données ?

<https://www.app.asso.fr/questions-frequentes/la-protection-des-bases-de-donnees/quels-sont-les-droits-reconnus-au-producteur-de-la-base-de-donnees.html>

L'*Open Access* (libre accès aux publications scientifiques) reste bien une simple faculté pour les chercheurs, qu'ils peuvent choisir ou non d'exercer en faisant jouer leur droit d'auteur. Mais tel n'est pas le cas pour l'*Open Data*, dans la mesure où les droits sur les bases de données de recherche appartiendront généralement aux établissements et non directement aux chercheurs.

II) L'ébauche d'un statut juridique applicable aux données de la recherche

a) Une protection contre la captation exclusive par les éditeurs privés

Comme nous l'avons vu plus haut, les données de recherche sont englobées par les principes généraux fixés par la loi Lemaire. Mais si le texte a fait disparaître le régime dérogatoire dont elles relevaient auparavant, il comporte à son article 30 quelques dispositions spécifiques aux données de recherche qui vont dans le sens de la protection du principe de libre réutilisation s'attachant désormais à elles (Maurel 2016¹¹) :

Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'État, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre.

L'éditeur d'un écrit scientifique mentionné au I ne peut limiter la réutilisation des données de la recherche rendues publiques dans le cadre de sa publication.

Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite.

Les deux derniers alinéas de cet article mettent en place un mécanisme qui va dorénavant protéger les données de recherche de la captation abusive par des éditeurs privés. Certains éditeurs avaient en effet commencé à développer des stratégies visant à demander aux chercheurs de leur remettre les données ayant servi à la réalisation d'un article scientifique en même temps que l'écrit associé. Souvent désignés par le terme de « *supplementary materials* », ces contenus pouvaient être soumis aux mêmes clauses d'exclusivité que les articles eux-mêmes par l'effet des contrats d'édition signés par les chercheurs. Le développement incontrôlé de ces pratiques aurait pu à terme conduire à une captation par les éditeurs des droits d'usage sur de grandes masses de données de recherche, avec à la clé des problèmes de concentration identiques à ceux que l'on rencontre actuellement à propos des articles scientifiques.

¹¹ <https://scinfolex.com/2016/11/03/quel-statut-pour-les-donnees-de-la-recherche-apres-la-loi-numerique/>

Le législateur français a néanmoins pris les devants en neutralisant par avance les cessions contractuelles de droits pouvant être imposés par les éditeurs. Même si les contrats d'édition comportent des clauses exclusives relatives aux données de recherche, celles-ci seront nulles d'effet et les chercheurs ne seront pas tenus de les respecter

b) La portée du principe de libre réutilisation des données de recherche

L'article 30 indique que lorsque les données issues d'une activité de recherche ont été « rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre ». Ce dernier terme mérite que l'on s'y attarde pour en préciser la portée.

Si on l'interprète comme signifiant « la réutilisation est complètement libre », cela voudrait dire que les données de recherche publiées seraient *ipso facto* placées dans le domaine public, déliant les réutilisateurs de toute obligation, y compris par exemple celle de citer la source des informations. Mais il est probable que le législateur n'ait pas entendu aller aussi loin. En effet, l'article L. 322-1 du Code des relations entre l'administration et le public (CRAP) fixe un certain nombre de conditions minimales à respecter en cas de réutilisation d'informations publiques :

Sauf accord de l'administration, la réutilisation des informations publiques est soumise à la condition que ces dernières ne soient pas altérées, que leur sens ne soit pas dénaturé et que leurs sources et la date de leur dernière mise à jour soient mentionnées.

Il est raisonnable de penser que ces conditions restent bien applicables aux données de recherche, même si elles ne sont pas reprises dans l'article 30. On notera que le « sauf accord de l'administration » permet néanmoins de les lever, au cas où un projet de recherche voudrait rendre complètement libre la réutilisation de ses données en les versant dans le domaine public.

c) Quelles licences pour la diffusion des données de recherche ?

L'*Open Data* s'exerce normalement dans le cadre de licences qui ont pour but de garantir aux réutilisateurs les droits dont ils disposent sur les données, tout en fixant certaines conditions. Durant la période antérieure à 2015, le développement de l'*Open Data* est passé en France par le recours à certaines licences, comme la licence ouverte/*Open Licence* élaborée par la mission Etalab pour les administrations centrales. Au niveau local, un éventail plus large de licences *Open Data* était utilisé par les collectivités locales (Jean 2011¹²).

Dorénavant, la situation va également changer de ce point de vue, car la libre réutilisation sera possible dès que des données publiques seront diffusées, même si leur producteur ne leur a pas associé de licence particulière. Ce sont directement les dispositions générales de la

¹² <https://commons.m.wikimedia.org/wiki/File:Jt-dans-la-jungle-des-licences.pdf>

loi qui entraîneront la libre réutilisation (sauf hypothèses exceptionnelles détaillées dans la partie suivante).

Le Code des relations entre le public et l'administration précise bien ce caractère désormais optionnel des licences de réutilisation :

La réutilisation d'informations publiques peut donner lieu à l'établissement d'une licence. Cette licence est obligatoire lorsque la réutilisation est soumise au paiement d'une redevance.

L'emploi du « peut » vaut en droit pour l'ouverture d'une simple faculté et en principe, les données de recherche ne peuvent désormais faire l'objet de redevances du fait du principe de gratuité posé par la loi Valter. Il en résulte que les projets de recherche resteront libres de choisir une licence ou non pour diffuser les données de recherche, sans que cela constitue une obligation. Cela signifie aussi que même si un projet n'attache pas de licence particulière à des données diffusées, elles n'en seront pas moins réutilisables. Au titre des bonnes pratiques, il paraît néanmoins utile de continuer à employer des licences de réutilisation, car celles-ci ont l'intérêt d'informer clairement les réutilisateurs sur l'étendue de leurs droits et obligations.

La loi ne laisse cependant pas toute latitude aux administrations pour choisir leur licence. Elle renvoie à un décret¹³ listant de manière limitative les licences pouvant être retenues pour encadrer la diffusion de données :

L'administration peut soumettre la réutilisation à titre gratuit des informations publiques qu'elle détient aux licences suivantes :

1° La licence ouverte de réutilisation d'informations publiques ;

2° L'Open Database License.

La licence ouverte est l'équivalent d'une licence *Open Source*, autorisant toutes les formes de réutilisation à la seule condition d'en citer la source. La licence ODbL comporte en plus une condition de « partage à l'identique » imposant aux utilisateurs de reverser sous la même licence les bases de données dérivées produites à partir des informations réutilisées.

Si une administration veut choisir une autre licence (ou en écrire une), elle ne pourra le faire qu'après avoir obtenu une homologation auprès de la direction interministérielle du numérique et du système d'information et de communication de l'Etat (DINSIC)¹⁴.

III) Les exceptions à la libre réutilisation des données de la recherche

¹³ Voir Décret n° 2017-638 du 27 avril 2017 relatif aux licences de réutilisation à titre gratuit des informations publiques et aux modalités de leur homologation : <https://www.legifrance.gouv.fr/eli/decret/2017/4/27/2017-638/jo/texte>

¹⁴ On notera que, de manière surprenante, les licences Creative Commons ne figurent pas dans cette liste, en dépit de leur réputation et du fait qu'elles peuvent être utilisées pour des bases de données depuis leur version 4. Leur utilisation dans le cadre d'un projet de recherche nécessitera donc une homologation au cas par cas.

a) L'articulation avec les droits de propriété intellectuelle

Le principe d'ouverture des informations publiques s'applique certes par défaut, mais la loi continue à fixer une série d'exceptions dans lesquelles il devra être écarté. Ces exceptions peuvent potentiellement concerner le secteur de la recherche et on peut même penser qu'elles trouveront davantage à s'y appliquer que dans le cadre d'autres activités administratives. C'est particulièrement le cas pour les droits de propriété intellectuelle. L'article L. 321-2 du Code des relations entre le public et l'administration prévoit en effet que :

Ne sont pas considérées comme des informations publiques, pour l'application du présent titre, les informations contenues dans des documents [...] sur lesquels des tiers détiennent des droits de propriété intellectuelle.

Or ce que l'on entend par « données de recherche » peut correspondre à des œuvres protégées par des droits de propriété intellectuelle appartenant à des tiers. Songeons par exemple à un projet qui entendrait procéder à de la fouille automatique de textes sur des articles de presse récents. De telles « données » ne pourraient pas être considérées comme des informations publiques et elles échapperaient donc au principe d'ouverture par défaut. Cela signifie même que de telles données ne doivent pas être mises en ligne et rendues librement réutilisables, sous peine de violer le droit d'auteur s'attachant à ces objets. Il conviendra donc de vérifier si les corpus de données à ouvrir ne correspondent pas à des œuvres toujours protégées, ce qui peut parfois s'avérer complexe à déterminer.

Par ailleurs, les chercheurs eux-mêmes sont considérés comme des tiers vis-à-vis de l'administration. Cette situation fait qu'ils restent pleinement titulaires de leurs droits d'auteur sur les écrits qu'ils produisent dans le cadre de leurs activités. Mais outre les articles et les ouvrages, les chercheurs peuvent aussi produire des « matériaux de recherche » assimilables à des œuvres protégées. Si l'on prend par exemple un projet de recherche en archéologie, nombre des pièces produites sur un chantier de fouille peuvent être protégées par le droit d'auteur (photos, dessins, plans, etc.). Bien que ces contenus puissent être qualifiés dans le langage courant de « données de recherche », ce ne sera pas le cas d'un point de vue légal. Les chercheurs disposeront de la faculté de décider s'ils veulent publier en ligne ces contenus protégés et les rendre librement réutilisables (sans pouvoir y être contraints).

Par ailleurs, la propriété intellectuelle pourra continuer à jouer sur des bases de données, si celles-ci sont produites par un établissement à caractère industriel et commercial (ce que sont certaines institutions de recherche : CNES, IFREMER, etc.). Nous avons expliqué plus haut que le droit de producteur de base de données avait été « neutralisé » par la loi République numérique, mais celle-ci apporte cette précision qui écarte les établissements publics à caractère industriel et commercial (EPIC) du principe d'ouverture par défaut :

Le premier alinéa du présent article n'est pas applicable aux bases de données produites ou reçues par les administrations [...] dans l'exercice d'une mission de service public à caractère industriel ou commercial soumise à la concurrence.

b) L'articulation avec la protection des données personnelles

L'ouverture des données publiques doit se conjuguer avec la protection des données à caractère personnel et de la vie privée. L'article 6 de la loi pour une République numérique est clair sur la nécessité de concilier ces deux impératifs :

Sauf dispositions législatives contraires ou si les personnes intéressées ont donné leur accord, lorsque les documents [...] comportent des données à caractère personnel, ils ne peuvent être rendus publics qu'après avoir fait l'objet d'un traitement permettant de rendre impossible l'identification de ces personnes.

Il ne suffit pas d'anonymiser des données pour satisfaire l'exigence de protection de la vie privée. Les données à caractère personnel peuvent en effet permettre la réidentification des personnes par recoupement, même sans les désigner nommément. Il n'est donc possible d'ouvrir des données de recherche et de les rendre librement réutilisables qu'à la condition de s'être assuré que l'identification des personnes est bien impossible (l'emploi de ce terme renvoyant à une obligation de résultat et non de moyens).

Là encore, il est plus que probable que les données de recherche soient assez fortement impactées par cette exception au principe de libre réutilisation. La sensibilité aux enjeux de protection de la vie privée varie selon les disciplines scientifiques, mais certains secteurs des sciences humaines et sociales (comme la sociologie, la psychologie, la géographie, etc.) appelleront une vigilance particulière.

Toutes les conditions posées par la loi informatiques et libertés s'appliquent en matière de traitement des données personnelles dans un cadre de recherche, sachant par ailleurs que ces exigences ont encore été renforcées par le Règlement général de protection des données (RGPD) entré en vigueur le mai 2018 12. Ce dernier fait peser une obligation de « redevabilité » (accountability) sur les utilisateurs de données personnelles, les obligeant à documenter leurs pratiques et à anticiper les risques. Il comporte un certain nombre de dérogations en faveur des activités de recherche destinées à faciliter les traitements, sans toutefois aller jusqu'à autoriser la rediffusion en Open Data de celles-ci¹⁵.

Ce nouveau cadre sera sans doute délicat à concilier avec le principe d'*Open Data* par défaut, surtout dans un domaine comme celui de la recherche où les informations sur les personnes peuvent s'avérer sensibles.

c) L'articulation avec la protection de la confidentialité et des secrets

¹⁵ Voir Lionel Maurel. Données personnelles et recherche : quelle articulation dans le RGPD ? S.I.Lex, 18 juillet 2018 [En ligne] : <https://scinfolex.com/2018/07/18/donnees-personnelles-et-recherche-scientifique-quelle-articulation-dans-le-rgpd/>

La loi République numérique établit un lien étroit entre le périmètre d'application du principe d'ouverture par défaut et celui des documents administratifs librement communicables.

Or la communication des documents peut être refusée au motif de protéger un certain nombre de secrets énumérés par la loi : secret défense, sûreté de l'État et sécurité publique, secret des délibérations du gouvernement, secret médical, etc¹⁶. S'il est possible qu'un chercheur obtienne une dérogation pour accéder à ce type d'informations dans le cadre de la conduite d'un projet de recherche, elles ne seront ensuite pas soumises au principe d'ouverture par défaut.

Le secret en matière commerciale et industrielle¹⁷ est aussi susceptible d'interférer avec l'ouverture par défaut des données de recherche. Celui-ci concerne normalement les administrations qui déploient leurs activités dans un domaine concurrentiel. Comme on l'a vu plus haut, certains établissements de recherche ont une nature industrielle et commerciale : ils pourront donc sans doute invoquer la protection de cette forme de secret pour échapper aux obligations de diffusion de données de recherche qu'ils produisent. Pour les universités et les établissements de recherche à caractère administratif, le recours au secret commercial paraît en revanche exclu, car les activités de recherche ne s'exercent pas en elles-mêmes dans un cadre commercial.

Le secret en matière commerciale et industrielle pourra aussi être invoqué par des entreprises privées, partenaires de projets de recherche, pour imposer la non-divulgence de certaines données susceptibles de fragiliser leur position sur un marché ou de révéler leurs procédés. Les accords de consortium signés pour le montage de ce type de partenariats public-privé de recherche comportent généralement des clauses de confidentialité et de partage de propriété sur les résultats de recherche, qui risquent d'interférer avec l'ouverture par défaut. Il ne s'agit pas cependant d'un blanc-seing laissé à la discrétion des partenaires : la confidentialité des données doit réellement permettre de protéger un secret, sous le contrôle de la CADA et *in fine* du juge administratif.

Conclusion : Le défi de la mise en pratique de l'ouverture des données de recherche

Les principes fixés par les lois Valter et Lemaire sont clairs : les données de recherche sont bien incluses dans le principe d'ouverture par défaut, tout comme le sont l'essentiel des informations produites par les autres administrations. Néanmoins, l'application concrète de ces règles risque de s'avérer plus complexe dans le secteur de la recherche que dans d'autres champs.

¹⁶ Voir CADA. Les secrets protégés : <http://www.cada.fr/les-secrets-protoges,58.html>

¹⁷ Voir CADA. Le secret en matière commerciale et industrielle : <http://www.cada.fr/le-secret-en-matiere-commerciale-et-industrielle,6239.html>

Nous avons vu en effet que les exceptions à la libre réutilisation sont susceptibles de s'appliquer fréquemment en matière de recherche, mais même en dehors de ces hypothèses, la recherche comporte par rapport à d'autres activités administratives des spécificités que la loi n'aborde pas en tant que telles. Par exemple, à quel moment doit intervenir la mise en ligne de données de recherche comprises dans le principe d'*Open Data* par défaut ? Faut-il procéder à cette diffusion dès que les données sont produites ou est-il permis d'attendre que les chercheurs aient effectué une première publication de leurs résultats pour leur en réserver la primeur ?

La loi ne règle pas en l'état ce type de problèmes qui renvoient moins à des questions juridiques qu'à la définition de la science ouverte (*Open Science*) et à la portée que l'on souhaite donner à ce concept. Les bonnes pratiques et l'éthique joueront dès lors sans doute un rôle aussi important que le droit pur dans le processus d'ouverture des données de recherche.

La position des financeurs de la recherche sera également déterminante. De ce point de vue, la Commission européenne a adopté depuis 2016 des lignes directrices intéressantes en matière d'ouverture des données de recherche. Dans le cadre du programme H2020, l'ouverture des données est désormais la règle, dans la mesure où la libre réutilisation est possible sans compromettre des droits de tiers¹⁸. Mais la Commission demande aussi que cette dimension de l'ouverture soit prise en compte en amont du déroulement des projets, dans le cadre de l'établissement d'un plan de gestion des données (Data Management Plan). Un guide pour la gestion des données de recherche a été mis à disposition des chercheurs qui recommande de suivre une méthodologie baptisée FAIR¹⁹ (pour *Findable, Accessible, Interoperable, Reusable*). Ce type d'approche pourrait être généralisée de manière à ce que l'ouverture des données devienne une dimension à part entière du processus de recherche, avec une prise en compte dès l'origine des questions juridiques.

La France s'est engagée dans une voie similaire avec la publication en juillet 2018 d'un Plan National pour la Science Ouverte prévoyant à l'un de ses points de "rendre obligatoire la diffusion ouverte des données de recherche issues de programmes financés par appels à projets sur fonds publics"²⁰. Cette politique a d'ores et déjà commencé à recevoir une traduction concrète, puisque l'Agence Nationale de la Recherche (ANR) a annoncé dans le cadre de son plan d'action 2019 sa volonté de favoriser l'ouverture en exigeant des candidats la réalisation d'un plan de gestion des données (DMP) pour les projets financés à partir de 2019²¹.

¹⁸ Voir Horizon 2020. Le libre accès aux publications et aux données de la recherche :

<http://www.horizon2020.gouv.fr/cid82025/le-libre-acces-aux-publications-aux-donnees-recherche.html>

¹⁹ Voir H2020 Programme. Guidelines on FAIR Data Management in Horizon 2020. Juillet 2016 :

http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

²⁰ MESRI. Plan National pour la Science Ouverte, 4 juillet 2018 [en ligne] : <http://m.enseignementsup-recherche.gouv.fr/cid132529/le-plan-national-pour-la-science-ouverte-les-resultats-de-la-recherche-scientifique-ouverts-a-tous-sans-entrave-sans-delai-sans-paiement.html>

²¹ ANR. Plan d'action 2019 [en ligne] : <http://www.agence-nationale-recherche.fr/PA2019#documents>

