



HAL
open science

Guide d'application de la Loi pour une République numérique pour les données de la recherche

Cécile Arènes, Lionel Maurel, Stephanie Rennes

► To cite this version:

Cécile Arènes, Lionel Maurel, Stephanie Rennes. Guide d'application de la Loi pour une République numérique pour les données de la recherche. Comité pour la science ouverte. 2022. hal-03968218

HAL Id: hal-03968218

<https://hal-lara.archives-ouvertes.fr/hal-03968218>

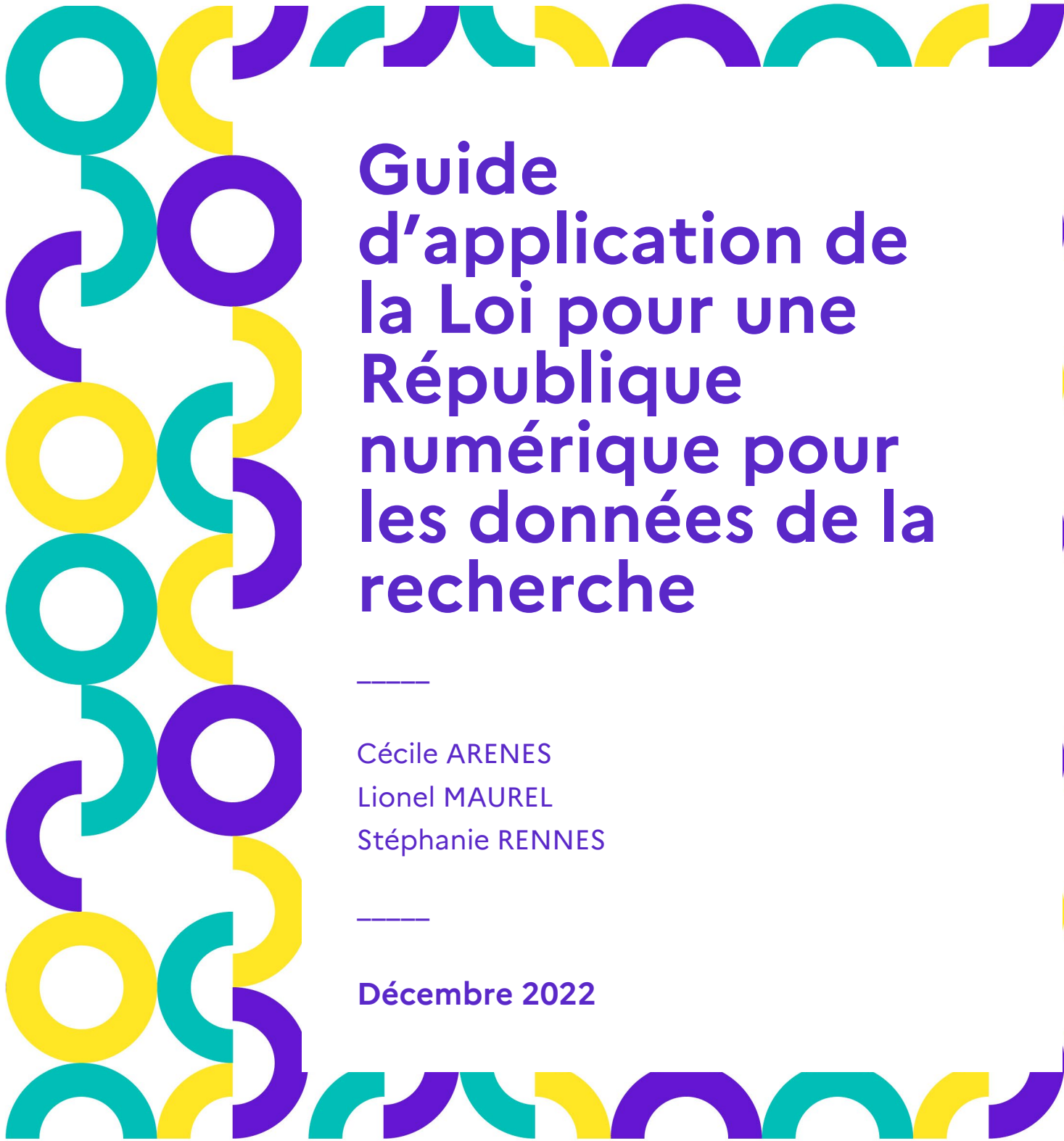
Submitted on 1 Feb 2023

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 - NoDerivatives | 4.0 International License



Guide d'application de la Loi pour une République numérique pour les données de la recherche

Cécile ARENES
Lionel MAUREL
Stéphanie RENNES

Décembre 2022

Guide d'application de la Loi pour une République numérique pour les données de la recherche

Collège Données de la recherche et Groupe d'expertise juridique du
Comité pour la science ouverte

Cécile ARENES
Sorbonne Université

Lionel MAUREL
CNRS

Stéphanie RENNES
Inrae

Décembre 2022

Conception graphique : opixido



Except where otherwise noted, this work is licensed under
<https://creativecommons.org/licenses/by-nd/4.0/deed.fr>

Sommaire

Introduction	4
Cadre général	5
Quel contexte de production des données ?	7
Exceptions à l'ouverture et à la libre réutilisation des données	8
Quand ouvrir les données ?	10
Rôle du chercheur et des établissements	10
Notion de données achevées	10
Qui peut réutiliser les données et pour quels usages ?	12
Licences	12
Pratiques des éditeurs et Loi pour une République numérique	14

Introduction

Promulguée en 2016, la loi pour une République numérique¹ (LRN) comporte plusieurs mesures sur lesquelles vient s'appuyer la politique nationale de science ouverte. Ces mesures sont transposées, notamment, dans le Code de la recherche et dans le Code régissant les relations entre le public et l'administration (CRPA).

Le présent guide vise à proposer des recommandations pour l'application de ces mesures.

En particulier, il propose des éléments d'interprétation de l'article 30 de la loi (devenu article L. 533-4 du Code de la recherche²), qui est connu pour être une disposition majeure en matière de publication des écrits scientifiques (alinéas I et III), en ce qu'il instaure un droit pour les chercheurs au dépôt de leurs publications en archive ouverte, moyennant le respect de certaines conditions. Cet article comporte également des précisions, peut-être moins connues, relatives aux données de recherche en tant que telles (alinéa II). **L'intérêt majeur de l'article L. 533-4 du Code de la recherche est qu'il vise des situations de production et de réutilisation des données scientifiques dans un contexte partenarial.**

La relation de cet article avec les dispositions du CRPA fixant le périmètre de l'« *open data*³ » des données publiques (principes et exceptions à leur ouverture et à leur réutilisation, cas particuliers, issus de l'article 6 de la LRN) est proposée comme fil directeur de ce guide.

Ce guide vient préciser les travaux antérieurs, notamment *Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France* de 2017⁴ et *Guide d'application de la loi pour une République numérique (article 30) Écrits scientifiques*⁵ de 2018, dans un contexte d'évolution constante de la réglementation, notamment au niveau européen.

¹ <https://www.legifrance.gouv.fr/loda/id/JORFTEXT000033202746>

² https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033205794

³ Définition par la CNIL : « L'*open data* désigne un mouvement, né en Grande-Bretagne et aux États-Unis, d'ouverture et de mise à disposition des données produites et collectées par les services publics (administrations, collectivités locales...) ».

⁴ <https://www.ouvrirelascience.fr/ouverture-des-donnees-de-recherche-guide-danalyse-du-cadre-juridique-en-france-v2/>

⁵ <https://www.ouvrirelascience.fr/guide-application-loi-republique-numerique-article-30-ecrits-scientifiques-version-courte/>

Cadre général

1. **Par principe, les données produites ou reçues dans le cadre de la recherche publique sont juridiquement considérées comme des « documents administratifs » ou des informations publiques.** Par conséquent, ce sont les établissements de recherche qui, en tant qu'administrations publiques, sont les garants, pour les collectifs de recherche, de la mise en œuvre de l'open data des données publiques et non les chercheurs à titre-individuel⁶.
2. **Les données de recherche issues de l'activité courante des établissements de recherche et détenues par eux sont ainsi soumises, sauf exceptions, aux principes d'ouverture par défaut et de libre réutilisation fixés par le CRPA.** Il faut noter que les données de recherche font également l'objet de dispositions complémentaires, prévues par l'article L.533-4 du Code de la recherche.
3. L'article L. 533-4 du Code de la recherche figure dans le chapitre consacré à "La valorisation des résultats de la recherche par les établissements et organismes de recherche". **Sa raison d'être n'est pas de fixer un régime dérogatoire pour les données de recherche (qui, lorsqu'elles correspondent à des données publiques, ont vocation à être ouvertes), mais d'apporter des précisions relatives aux situations où ces données sont produites dans le cadre de projets partenariaux associant des financements publics et des financements privés.**

Cette interprétation s'appuie notamment sur les orientations du rapport de la mission Bothorel, rappelées dans l'extrait ci-dessous :

⁶ Voir la section « rôle du chercheur et des établissements »

« Les données de la recherche : un régime spécial dans un domaine alliant données et codes sources publics et privés »

La recherche est un domaine à la frontière du public et du privé. Si une donnée est produite par un institut de l'enseignement supérieur et de la recherche, alors cette donnée est publique du fait du caractère administratif de l'institut (...). Dans ce cadre, le régime juridique décrit plus haut s'applique.

Dans le cas où une donnée est produite par un acteur privé ou ne relevant pas directement d'une administration, le critère du financement permet de déterminer si une donnée est publique ou non. Dans ce cas, la communication de la donnée suit les règles décrites plus haut, et sa réutilisation est libre si elles ont été publiées par l'organisme ou par le chercheur. »

Extrait du Rapport de la mission Bothorel⁷ - Pour une politique publique de la donnée, 2020 (p. 64), faisant respectivement référence aux dispositions du CRPA et du Code de la recherche.

⁷https://www.gouvernement.fr/upload/media/default/0001/01/2020_12_rapport_-_pour_une_politique_publicque_de_la_donnee_-_23.12.2020__0.pdf

Quel contexte de production des données ?

D'après l'article 533-4-II du code de la recherche, sont concernées « les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'Etat, des collectivités territoriales, des établissements publics, des subventions d'agences de financement nationales ou par des fonds de l'Union européenne (...) ».

1. Principal critère de l'exercice d'une « activité de recherche financée au moins pour moitié par des dotations » publiques, nationales ou européennes.

Ceci correspond, dans les faits, à la situation, prévue par le CRPA, de production d'informations publiques dans le cadre d'une mission de service public, déclenchant l'obligation d'Open Data pour les administrations. Ainsi, **sauf exceptions, les établissements d'enseignement supérieur et de recherche sont tenus de publier spontanément en ligne et de rendre librement réutilisables les informations publiques qu'elles produisent ou reçoivent⁸ dans le cadre de leur mission de service public.**

Pour un établissement de recherche, l'article 533-4-II du code de la recherche permet de préciser, dans un **contexte de recherche partenariale** soutenue à la fois par des fonds publics et privés, à quelles conditions les données de recherche doivent être considérées comme des données publiques.

A contrario, on doit lire l'article 533-4 comme signifiant que des données issues d'une recherche financées à plus de 50% par des acteurs privés sont exclues du principe d'ouverture par défaut et de libre réutilisation fixés aux articles L. 311 s. et L.312 s. du CRPA.

Il est proposé ici d'étendre aux données l'interprétation déjà recommandée par le Comité pour la science ouverte en matière de publications, à savoir que l'on doit prendre en compte les coûts complets de la recherche pour évaluer les 50% de financements publics, et notamment les salaires, et pas uniquement les subventions versées aux projets.

« La loi s'applique pour les auteurs, quel que soit leur statut (chercheur, ingénieur, doctorant, ... relevant d'un statut public ou privé) et leur nationalité, d'un écrit scientifique issu d'une activité de recherche financée au moins pour moitié par des dotations de l'État français, des collectivités territoriales ou des établissements publics, par des subventions d'agences de financement nationales ou par des fonds de l'Union européenne. L'évaluation du financement inclut tous les coûts de la recherche (salaires, hébergement ...). »

La combinaison des dispositions du CRPA et du code de la recherche amène à constater que, **lorsque les conditions sont réunies, les données de recherche liées à des publications – au même titre que toutes les données de recherche – sont des données publiques soumises aux principes d'ouverture par défaut et de réutilisation libre et gratuite, sauf si elles relèvent d'une des exceptions listées par la loi** (voir ci-dessous).

⁸ Le principal élément déclencheur de l'open data est la détention des données par l'établissement public, que ce soit du fait d'une production ou suite à leur réception par cet établissement (voir par exemple l'article L.300-2)

2. Dans le cadre partenarial auquel s'applique l'article 533-4-II du code de la recherche, les modalités de partage figurent dans les clauses contractuelles convenues entre les partenaires de recherche.

Ces clauses doivent respecter **deux conditions cumulatives** énoncées et commentées ci-après :

« Dès lors que les données issues d'une activité de recherche (...) 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 ».

Outre le critère du seuil de dotation publique du projet de recherche, dont on a vu qu'il permet de déterminer si l'on est potentiellement en présence de données publiques, il faut également **vérifier, avant toute forme de partage, l'existence d'un « droit spécifique » ou une « réglementation particulière »**. En requérant, par ailleurs, une mise à disposition publique préalable par « le chercheur, l'établissement ou l'organisme de recherche », le code de la recherche demeure en cohérence avec le CRPA. Ces dispositions, comme nous le rappelions en introduction, sont toutes deux issues de la même Loi pour une République numérique.

« Les dispositions du présent article sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite »⁹

Enfin, **le contrat ne doit pas contrevenir à l'ordre public** : le contrat, qui exprime la volonté des partenaires pour atteindre un objectif donné, ne peut aller à l'encontre de principes fondamentaux justifiés par des intérêts collectifs supérieurs, tels que l'exercice de droits d'accès consacrés par les textes (open data, accès à l'information environnementale ...) ou la protection de libertés fondamentales (données personnelles ...). Les dispositions de l'article ont également une portée que le législateur a élevée au rang d'ordre public. L'« ordre public » ne peut être contourné par les clauses contractuelles, sous peine de n'avoir aucun effet juridique pour les partenaires et de paralyser les recherches concernées.

Exceptions à l'ouverture et à la libre réutilisation des données

L'article 533-4-II du code de la recherche cite les « *Données protégées par un droit spécifique ou une réglementation particulière* » qui correspondent à des exceptions au principe d'ouverture énoncés fixés dans le CRPA.

Ces données visées par une exception sont les suivantes :

- Les données non communicables au sens du CRPA, par exemple celles dont la consultation ou la diffusion porterait atteinte au secret de la défense nationale, au secret des délibérations du gouvernement, à la conduite de la politique extérieure de la France, à la sûreté de l'Etat, à la sécurité publique, à la sécurité

⁹ Voir : Guide d'application de la loi pour une République numérique (art. 30) – Ecrits scientifiques <https://www.ouvrirlascience.fr/guide-application-loi-republique-numerique-article-30-ecrits-scientifiques-version-courte/>

des personnes, à la sécurité des systèmes d'information des administrations ou encore à « d'autres secrets protégés par la loi ».

- Les données communicables seulement à l'intéressé (vie privée, secret médical ; pour les entreprises du secteur privé, notamment, le secret des affaires : secret des procédés, des informations économiques et financières et des stratégies commerciales ou industrielles...).
- Les documents réalisés en exécution d'un contrat de prestation de services exécuté pour le compte d'une ou de plusieurs personnes déterminées.

D'autres types de données peuvent quant à elles être protégées par « une réglementation particulière » ou par « des droits spécifiques », par exemple :

- Les données correspondant à des informations à caractère personnel, qui ne peuvent être rendues réutilisables qu'après anonymisation ou avec le consentement des personnes concernées ou sur la base d'une autorisation légale ou réglementaire.
- Les données couvertes par des droits de propriété intellectuelle appartenant à des tiers : il s'agit d'hypothèses où les données correspondent à des œuvres de l'esprit originales, protégées à ce titre par le droit d'auteur.

Les pratiques de partage des données en accès restreint peuvent être favorisées à travers la définition de protocoles maîtrisés. Pour certaines données confidentielles, des formes de partage sécurisé peuvent par exemple être envisagées. Le RGPD prévoit notamment des dérogations en faveur des activités de recherche qui autorisent, sous certaines conditions encadrées, à partager des données personnelles entre chercheurs¹⁰. Dans le cadre de partenariats avec des entreprises, le partage de données couvertes par le secret commercial ou la propriété intellectuelle du partenaire nécessite leur accord.

Les données relevant de la protection du potentiel scientifique et technique de la nation (notamment certaines données issues des laboratoires situés en zone à régime restrictif – ZRR) constituent un cas particulier. Toutes les données produites par des laboratoires situés en zone à régime restrictif ne sont pas automatiquement exclues du principe d'ouverture par défaut. Pour déterminer les données à garder confidentielles, il convient de se rapprocher des personnes habilitées à se prononcer sur les restrictions de diffusion (fonctionnaire sécurité et défense de l'établissement, par exemple). Ensuite, comme dans tout autre projet, il revient aux équipes de partager les données publiques et achevées qui peuvent être ouvertes.

¹⁰ Voir en ce sens le décret n° 2018-687 du 1er août 2018 pris pour l'application de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés, modifiée par la loi n° 2018-493 du 20 juin 2018 relative à la protection des données personnelles :

« Les données issues de ces traitements conservées par le responsable du traitement ou son sous-traitant ne sont accessibles ou modifiables que par des personnes autorisées. Ces personnes respectent les règles de déontologie applicables à leurs secteurs d'activités. Les autorisations accordées par les responsables de traitement à ces personnes respectent les finalités spécifiques de l'alinéa précédent ainsi que les garanties prévues à l'alinéa suivant. »

Quand ouvrir les données ?

Rôle du chercheur et des établissements

Selon l'article 533-4-II du code de la recherche, lorsque les données « *ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, la réutilisation est libre.* ». Cette phrase suppose que l'on a suivi, a minima, le cheminement habituel vers l'open data, conformément aux dispositions du CRPA et du code de la recherche, que les exceptions à l'ouverture ont préalablement été écartées et que l'existence de droits spécifiques ou de réglementations particulières a été prise en compte. Des politiques de gouvernance ou de gestion des données de la recherche sont progressivement mises en place par un grand nombre d'établissements pour faciliter cette étape. Cet article 533-4-II ne doit donc pas être lu comme instaurant au bénéfice des chercheurs ou des établissements une faculté d'apprécier discrétionnairement si les données qu'ils produisent sont rendues librement réutilisables.

Dans un contexte partenarial, le comité de pilotage du projet est le lieu privilégié de l'examen de ces questions, avec l'aide des services d'appui (administrateurs des données, des algorithmes et des codes sources des établissements (ADAC), services de valorisation, services juridiques, délégués à la protection des données, fonctionnaire sécurité défense, documentalistes...). La rédaction des plans de gestion de données doit constituer par ailleurs une occasion d'envisager les questions de mise à disposition et de réutilisation dès l'amont du projet.

A noter que les établissements de recherche qui ne mettent pas spontanément en ligne les données couvertes par le principe d'ouverture peuvent se voir adresser des demandes de communication et de réutilisation des données (par le biais de procédures d'accès aux documents administratifs, notamment) auxquelles ils seront juridiquement tenus de faire droit.

Notion de données achevées

D'après le CRPA (article L. 311-2), l'application du principe d'ouverture des données publiques est liée à la notion de « **document achevé** ».

La question qui se pose alors est de savoir quand les chercheurs et établissements sont tenus de rendre publiques les données de recherche ou, à défaut d'une mise à disposition spontanée, quand le principe de libre réutilisation leur devient opposable.

Savoir exactement quand des données de recherche doivent être considérées comme « achevées » n'est pas une question simple à traiter et il est difficile de donner une réponse dans l'absolu. Ce caractère d'achèvement doit vraisemblablement s'apprécier dans chaque situation particulière et devrait faire l'objet d'une discussion au sein des communautés de recherche afin d'établir des standards en la matière, ayant un sens dans chaque discipline.

Toutefois, on peut considérer que les données sous-jacentes à une publication présentent un caractère achevé. Même si les données sont susceptibles de continuer à évoluer ensuite, elles présentent au moment de la publication un degré d'achèvement suffisant pour servir de fondement à ces résultats de recherche présentés devant la communauté scientifique. S'il peut être admis qu'un délai de mise à disposition soit respecté après la production de données de recherche pour permettre leur exploitation, **il ne devrait pas être possible de maintenir un « embargo » sur l'accès aux données sous-jacentes à un écrit scientifique après sa publication (lorsqu'elles correspondent à des données publiques).**

A noter que l'interprétation proposée ci-dessus ne signifie pas que les données de recherche doivent être considérées comme achevées uniquement si elles ont fait l'objet d'une publication scientifique. **D'autres éléments peuvent entrer en ligne de compte pour établir un degré d'achèvement suffisant (fin d'un projet de recherche, délai admis dans une communauté de recherche, etc.).**

Qui peut réutiliser les données et pour quels usages ?

Une fois les conditions précitées réunies, l'article 533-4-II du code de la recherche dit des données que « *Leur réutilisation est libre* ».

Ce passage est un **rappel du principe plus général de libre réutilisation des données publiques par le public qui figure dans le CRPA**. La réutilisation des données ne se fait pas pour autant sans quelques conditions à respecter listées à l'article L. 322-1 du CRPA :

« *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.* »¹¹

On doit donc considérer que, pour toute personne, **la réutilisation est libre, à condition de respecter les conditions minimales fixées par cet article : mention de source, non altération et non dénaturation, mention de la date de mise à jour. Ces obligations constituent des garanties fortes pour les chercheurs et le respect de l'intégrité des travaux de recherche.**

Licences

Selon l'article 533-4 du code de la recherche, la réutilisation des données publiques de recherche est libre, dès lors que les données sont publiées et n'entrent pas dans l'une des exceptions précitées et ce, même si aucune licence de réutilisation n'est associée aux données. Le CRPA est clair à ce sujet : les licences sont seulement optionnelles¹² et ne servent qu'à indiquer plus précisément leurs droits aux utilisateurs. En leur absence,

¹¹ Voir :

https://www.legifrance.gouv.fr/codes/section_lc/LEGITEXT000031366350/LEGISCTA000032255218/#LEGISCTA000032255218

¹² Les licences ne sont obligatoires que lorsque la réutilisation est soumise au paiement d'une redevance. Article L323-1 du CRPA : « *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.* »

Art. D. 323-2-1.-I.-L'administration peut soumettre la réutilisation à titre gratuit des informations publiques qu'elle détient aux licences suivantes.

c'est le principe de libre réutilisation¹³ de la loi qui prend le relais. Par ailleurs, des données publiques publiées sur un site (privé ou public) avec des conditions d'utilisation contraires au principe de libre réutilisation restent librement réutilisables.

Le CRPA n'impose pas une obligation d'attribuer une licence aux données pour la réutilisation par des tiers. Néanmoins, le recours aux licences de réutilisation est vivement recommandé par les financeurs de la recherche car c'est un moyen efficace de se conformer aux principes FAIR¹⁴, selon lesquels :

- *R1 Les données et les métadonnées ont des attributs multiples et pertinents.*
- *R1.1 Les données et les métadonnées sont mises à disposition selon une licence explicite et accessible.*
- *R1.2 Les données et les métadonnées¹⁵ sont associées à leur provenance.*
- *R1.3 Les données et les métadonnées correspondent aux standards des communautés indiquées.*

Dans l'acronyme FAIR, R-« REUSABLE » pose le principe d'une diffusion des données avec une licence d'utilisation, dans un objectif de traçabilité et de bon usage.

Si une licence est choisie pour diffuser des données de recherche, elle doit l'être parmi une liste fermée fixée par décret :

<https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000034502557/>

Les licences Creative Commons, largement utilisées à l'international et recommandées par les financeurs, ne figurent pas pour l'instant dans ce décret, dont une modification est à l'étude pour les intégrer. Il est néanmoins possible d'attribuer une double licence, par exemple Licence Ouverte française et CC:BY, à un jeu de données.

¹³ A savoir une réutilisation par le public, à des fins autres que celles pour lesquelles les données ont été produites, en citant le producteur (source), en indiquant la date de la dernière mise à jour des données, avec interdiction d'altérer ou de dénaturer ces dernières.

¹⁴ Les [principes FAIR](#) (Findable, Accessible, Interoperable, Reusable) décrivent comment les données doivent être organisées pour être plus facilement accessibles, comprises, échangeables et réutilisables.

¹⁵ Littéralement, une [métadonnée](#) (metadata) est une donnée sur une donnée. Les métadonnées peuvent aussi se définir comme un ensemble structuré de données qui servent à définir ou décrire une ressource quel que soit son support (papier ou électronique).

Pratiques des éditeurs et Loi pour une République numérique

« 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. », article 533-4-III du code de la recherche

Cet article instaure un mécanisme de protection pour les données de recherche liées à une publication scientifique. On peut vraisemblablement lui donner deux significations différentes (qui se cumulent) :

1. **Les éditeurs d'écrits scientifiques ne peuvent valablement obtenir de cessions exclusives de droits sur des données de recherche liées à la publication, qu'elles soient déposées sur un entrepôt de données ou qu'elles figurent comme *supplementary materials* de l'article.**
2. **Lorsque des éditeurs rassemblent des données de recherche liés à des publications dans des entrepôts, ils ne peuvent faire jouer leur droit de producteur de bases de données pour limiter la réutilisation.**

Ce passage de l'article 533-4-III du code de la recherche constitue une forte protection des données de recherche. Il garantit que les éditeurs ne peuvent exiger d'exclusivité et donc empêcher que les données leur ayant été fournies soient diffusées par ailleurs par les chercheurs. Les éditeurs ou les revues peuvent néanmoins exiger le dépôt des données liées aux publications dans leur entrepôt ou sur leur plateforme. Mais dans ce cas, les éditeurs ne peuvent empêcher la libre réutilisation des données par le biais de conditions d'utilisation. C'est le sens de l'article 533-4-IV qui dispose que "*les dispositions [de l'article 533-4] sont d'ordre public et toute clause contraire à celles-ci est réputée non écrite*". Cette affirmation concerne donc également les contrats signés avec des éditeurs étrangers.

Malgré cette protection conférée par la loi, il reste néanmoins vivement recommandé aux chercheurs de déposer les données de recherche dans des entrepôts répondant à des critères de confiance, maîtrisés par des acteurs publics et les communautés de recherche (entrepôts institutionnels, infrastructures nationales ou entrepôts disciplinaires ou entrepôt national pluridisciplinaire Recherche Data Gouv) et cela en amont de la publication.