

L'ouverture des données publiques : la République numérique entre dans le monde de la recherche

Nathalie Gandon¹, Nathalie Morcrette²

Résumé. La réglementation française est à la pointe en matière d'ouverture des données ou « open data » car c'est l'une des plus incitatives au monde. Poussé par le désir de transparence de l'action publique, le législateur français a imposé à toutes les composantes de l'État la diffusion, sous un mode libre et gratuit, des données qu'elles produisent. La loi pour une « République numérique » d'octobre 2016 se traduit par une obligation générale de diffusion à tout public des données, y compris celles des organismes publics d'enseignement supérieur et de recherche. Mais le principe général de diffusion masque la complexité des exceptions à la diffusion. Ces exceptions sont nombreuses et doivent être passées en revue avant la mise en ligne des jeux de données. Pour les équipes de recherche, l'obligation est donc à la fois une contrainte car elle amène à s'interroger sur la communicabilité des données au public et une opportunité car le nombre de jeux de données en libre accès est voué à augmenter rapidement, permettant d'interroger et de croiser de nouvelles données. Pour se retrouver dans les méandres réglementaires de l'open data, un groupe de spécialistes a produit un guide d'analyse du cadre juridique de l'ouverture des données de recherche. Le présent article en propose un condensé.

Mots clés : ouverture des données, open data, République numérique, données de la recherche, diffusion

Introduction

La loi pour une République numérique, très médiatisée, du fait notamment de son processus d'élaboration qui a fait appel au public, a mis en lumière la numérisation de notre environnement quotidien, personnel et professionnel. Cette numérisation concerne également l'État, sans cesse invité à se moderniser et à rendre accessibles les informations qu'il possède. Le but étant de permettre aux citoyens et aux acteurs économiques d'apprécier l'action de l'État et d'en tirer des fruits. Les acteurs publics de la recherche ne sont pas à côté de ce mouvement de « digitalisation » et d'ouverture. Nous verrons qu'ils sont directement concernés.

Pour se retrouver dans les méandres réglementaires de l'open data, un groupe de spécialistes a produit un guide d'analyse du cadre juridique de l'ouverture des données de recherche (Becard et al., 2017). Le présent article en propose un condensé.

Nous allons donc essayer de comprendre ce que signifie l'ouverture des données publiques pour les données de recherche d'un point de vue juridique, ses contraintes, ses opportunités. La réglementation concernant le libre accès aux publications scientifiques (open access) n'est pas abordée, ce sujet étant complémentaire mais indépendant de l'open data, qui concerne les données complémentaires ou indépendantes des publications.

¹CODIR-MIA, Collège de Direction, Département MIA, Inra, Paris
Nathalie.gandon@inra.fr

²DPTI, Direction du Partenariat et du Transfert pour l'Innovation, Inra, Paris
Nathalie.morcrette@inra.fr

Motivations économiques et politiques de l'open data

En préambule, rappelons en quelques traits les **objectifs et motivations** de l'ouverture des données publiques.

Tout d'abord, cette ouverture n'est pas un concept nouveau. Déjà, la déclaration des droits de l'homme et du citoyen (26.08.1789) avait édicté un principe de **transparence** de l'action de l'administration. En effet, cette déclaration mentionne que « *La société a le droit de demander compte à tout agent public de son administration* ». Ce concept de transparence de l'action de l'État peut être « lu dans les deux sens ». En effet, l'État rend compte au citoyen de son action, le citoyen à son tour, propose à l'État de se transformer, s'améliorer. Ainsi, la transparence de l'État et le droit de participation du citoyen vont de pair : la société a besoin de connaître la façon dont l'État est régulé, géré, pour pouvoir faire valoir son avis ; si le citoyen a demandé à l'État de se transformer ou a fait part de dysfonctionnements, il doit pouvoir ensuite juger sur « pièces » (en accédant aux données générées par les actions de l'État).

Plus récemment, l'Etat a vu dans la numérisation et dans l'ensemble des services qui peuvent être proposés autour de celle-ci une solution possible aux crises de l'emploi et de la croissance économique. Ainsi, la mise à disposition par l'Etat de ses données est vue comme une source potentielle de **développement d'innovations** et partant, de créations d'emplois. L'impact économique potentiel de l'ouverture des données de l'Etat est ainsi devenu le moteur de la levée de des barrières juridiques et techniques encore existantes avant de parvenir à l'open data.

Les principes retenus pour l'ouverture généralisée des données de l'Etat sont les suivants :

- le libre accès aux données ;
- une réutilisation des données possible par tous ;
- une absence de restrictions juridiques, techniques et financières à l'accès et à la réutilisation.

Par ailleurs, les principes scientifiques qui sous-tendent la qualité de l'open data sont résumés par le standard « FAIR » : findable, accessible interoperable, reusable, autrement dit les données doivent être (re)trouvables, accessibles (donc ouvertes), interoperables, réutilisables.

Principaux textes régissant l'open data

Venons-en maintenant à la **réglementation** et aux textes principaux régissant l'open data.

Nous ne mentionnerons pas l'exhaustivité des textes, ce qui serait quasiment impossible. Car beaucoup de nouvelles lois françaises comportent désormais un article obligeant à rendre publiques tel ou tel type de données (exemples : données de transport, données des conventions de subvention).

Tout d'abord, il est important de rappeler le contexte réglementaire européen. C'est celui de la directive « PSI » ou « public sector information » qui date de 2003 et a été revisitée en 2013. Ce texte concerne la réutilisation des informations du secteur public et oblige à une mise à disposition par les Etats européens de leurs données administratives.

Il est intéressant de noter que sont notamment exclus du champ d'application de cette directive les documents :

- issus des établissements d'enseignement et de recherche sauf les documents détenus par les bibliothèques universitaires pour lesquels existe un régime spécifique ;
- sur lesquels des tiers détiennent des droit de propriété intellectuelle.

Cette directive est transposée en droit français. Toutefois, **le législateur français a décidé d'aller plus loin que le texte européen**. En conséquence, les obligations d'open data qui s'imposent aux établissements français d'enseignement supérieur et de recherche n'ont pas forcément d'équivalent pour nos partenaires européens. Ces différences de législations sont néanmoins « gommées » dans la pratique par les exigences des financeurs (publics), qui demandent de plus en plus souvent à ce que les données des recherches qu'ils ont financées soient rendues publiques.

Nathalie Gandon, Nathalie Morcrette

Il faut tout d'abord souligner que la réglementation nationale est bien plus ancienne que la réglementation européenne puisque c'est la loi dite « CADA » (commission d'accès aux documents administratifs) de 1978 qui a instauré un principe d'accès aux documents produits ou détenus par l'Etat.

Depuis quelques années, cette loi a été codifiée dans le code des relations entre le public et l'administration (ou « CRPA »). Elle a fait l'objet récemment de deux évolutions majeures :

- en décembre 2015 : la loi « Valter » est venue supprimer la faculté pour les établissements de recherche et d'enseignement supérieur de fixer librement les modalités de réutilisation des documents administratifs qu'ils émettent et a instauré un principe de gratuité de la réutilisation des informations publiques (sauf exceptions très limitatives) ;
- en octobre 2016 : la loi pour une République numérique (ou loi « Lemaire ») a inversé la logique de diffusion des documents administratifs : l'État ne doit plus attendre la demande du citoyen pour communiquer ses documents mais doit diffuser les informations de manière volontaire. De plus, les codes source des logiciels produits par l'administration sont désormais considérés comme des documents administratifs, soumis à obligation de diffusion au public.

Notion d'open data dans la réglementation française

L'élargissement de la notion de document administratif par les textes récents nous amène à nous poser la question suivante : quelles sont les productions des établissements d'enseignement supérieur et de recherche qui peuvent être considérés comme des documents administratifs ?

Selon l'article L300-2 du CRPA : « *Sont considérés comme documents administratifs, que soient leur date, leur lieu de conservation, leur forme et leur support, les documents produits ou reçus, dans le cadre de leur mission de service public, par l'Etat, les collectivités territoriales ainsi que par les autres personnes de droit public ou les personnes de droit privé chargées d'une telle mission. Constituent de tels documents notamment les dossiers, rapports, études, comptes rendus, procès-verbaux, statistiques, instructions, circulaires, notes et réponses ministérielles, correspondances, avis, prévisions, codes sources et décisions.* ».

Les établissements publics à caractère scientifique et technologique (EPST) tels que l'Inra entrent donc bien dans le champ de cette disposition : ils produisent des documents administratifs, qu'ils doivent rendre publics.

La loi parle de « documents administratifs » mais ajoute parfois la notion « d'informations publiques ». Ces notions sont souvent équivalentes pour la compréhension de la réglementation car les informations publiques ne sont autres que le contenu d'un document administratif, autrement dit les données.

Lorsque des rapports, bases de données, ou autres productions sont qualifiés de documents administratifs, les conséquences suivantes s'appliquent :

- les documents administratifs sont susceptibles d'**accès**, autrement dit le citoyen peut demander communication à l'administration de tout document administratif (nous verrons plus loin que ce principe souffre des exceptions) ;
- les documents administratifs sont susceptibles de **réutilisation**. En principe, ce qui est accessible est également réutilisable. Cependant ce principe souffre des exceptions. Ainsi, par exemple, pour des données relevant du secret médical : des médecins peuvent y avoir accès mais ne peuvent pas les divulguer (réutiliser).

Au-delà des caractéristiques d'accès et de réutilisation, l'application aux documents administratifs du concept d'open data tel que revisité par les lois françaises récentes fait entrer l'information publique dans des dimensions d'universalité et de gratuité.

- **Universalité** ? Oui, car l'information publique devient réutilisable par tous et pour tous les usages. Ainsi, selon l'article. L321-1 CRPA : « *Les informations publiques figurant dans des documents communiqués ou publiés par les administrations [...] peuvent être utilisées par toute personne qui le souhaite à d'autres*

fins que celles de la mission de service public pour les besoins de laquelle les documents ont été produits ou reçus. ». Cette notion d'universalité de la réutilisation concerne aussi le format des informations publiques diffusées : celles-ci doivent être accessible dans un format standard et aisément réutilisable (les formats de fichiers bloquant sont proscrits).

- **Gratuité ?** Oui, car la réutilisation de l'information publique doit être gratuite. L'application d'une redevance qui pouvait être une pratique courante pour certains établissements publics avant la loi Valter (comme, par exemple, pour Météo France) devient désormais l'exception. Le prix doit maintenant être justifié par des considérations économiques, autorisé par décret au terme d'une procédure d'agrément et révisé tous les 5 ans.

En corollaire de ces notions, pour parfaire l'ouverture des données à laquelle doivent se conformer toutes les composantes de l'État, les lois récentes imposent ou réaffirment :

- la nécessité de mettre en place un répertoire des informations publiques dans chaque établissement ;
- l'obligation de diffusion spontanée des informations disponibles sous format électronique ;
- la diffusion d'une liste des bases de données produites par l'administration mises à jour régulièrement ;
- la publication des «...données, mises à jour de façon régulière, dont la publication présente un intérêt économique, social, sanitaire ou environnemental ... » (Art. L312-1-1 CRPA).

Application de l'open data aux productions issues d'un établissement de recherche

Avant de diffuser des données de recherche en open data, un certain nombre de vérifications s'impose. Nous proposons ci-dessous une méthodologie progressive, en quatre temps.

Il s'agit d'abord de s'intéresser à (i) la nature du résultat concerné, (ii) l'(les) auteur(s) du résultat et (iii) les conditions de production du résultat. Puis il faudra vérifier si l'une des nombreuses exceptions à l'open data s'applique, empêchant ainsi la diffusion de tout ou partie des informations.

Temps 1 : la nature de la production (du résultat)

La première vérification à réaliser tient à la nature du résultat à divulguer. Bien que la définition de document administratif soit très large, elle ne couvre que les documents existants sous forme papier ou numérique. On peut donc en déduire que le matériel biologique ainsi que les équipements (prototype, etc.) sont exclus de l'open data. Les logiciels et bases de données sont, quant à eux, inclus. Plus encore, pour ce qui concerne les données, la réglementation ne fait pas la distinction entre données brutes, données élaborées, métadonnées : toutes ces données entrent donc dans le champ de l'open data.

En revanche, sauf pour les informations environnementales, **les documents susceptibles d'être diffusés au public doivent être achevés**. Cette condition est très importante, surtout pour des productions issues de la recherche. Ainsi, les cahiers de laboratoires, les documents préliminaires à des articles scientifiques ne seraient pas communicables.

Temps 2 : l'(les) auteur(s)

Sans ambiguïté, les documents produits par les personnels administratifs et techniques de l'établissement dans le cadre de leurs fonctions sont des documents administratifs.

Cas particulier de la recherche : la situation est plus délicate pour ce qui concerne les documents produits par les chercheurs, et plus généralement par les personnels disposant d'une autonomie (c'est-à-dire les auteurs d'œuvres dont la divulgation n'est soumise à aucun contrôle préalable de l'autorité hiérarchique selon le code de la propriété intellectuelle). En effet, **ces personnels sont titulaires des droits d'auteur sur leurs productions**. En conséquence, les productions originales (c'est-à-dire marquées par l'empreinte de l'auteur) que sont ou peuvent être les articles scientifiques, les cartographies, les dessins, les photographies, qui sont réalisées par des chercheurs ne sont pas soumises à l'obligation d'open data. Elles peuvent néanmoins faire l'objet d'une diffusion sous licence libre et gratuite, autrement dit entrer dans un régime d'open access.

Revenons à la règle générale : les productions qui ne sont pas soumises au droit d'auteur, ce qui est souvent le cas des bases de données, appartiennent en principe à l'Institut et entrent pleinement dans le champ de l'open data. C'est le cas également des logiciels qui bénéficient d'une exception au droit d'auteur et entrent également pleinement dans le champ de l'open data.

Si l'on a vu que les choses sont claires pour toutes les productions réalisées par des agents, des questions sont à se poser pour les productions réalisées par des stagiaires ou doctorants. **Pour les stagiaires**, la signature d'une convention de cession de droits est le plus souvent nécessaire pour pouvoir disposer de tous les droits sur ses productions (écrits, logiciels, etc.) avant la mise en open data. **Pour les doctorants**, si le doctorant n'est pas un salarié de l'Institut, il convient de se reporter au contrat de recherche ou au contrat d'accueil pour savoir à qui appartiennent les résultats qu'il produit. En l'absence de contrat, il faudra trouver un accord, à matérialiser par écrit.

Temps 3 : les conditions de production du résultat

Lorsque le résultat que l'on souhaite diffuser en open data a été produit dans le cadre d'une Unité de recherche, mixte ou non, sans collaboration avec un tiers, il s'agit bien d'un document administratif, diffusable en open data.

Lorsque le résultat a été produit dans le cadre d'une collaboration avec un établissement public ou avec un établissement privé mais dans le cadre d'une mission de service public (exemple : résultats acquis avec la SNCF Réseau dans le cadre de sa mission d'entretien du réseau ferré), il s'agit là encore d'un document administratif.

Lorsque le résultat a été produit en collaboration avec un partenaire privé pour ses besoins (et en dehors d'une mission de service public), la qualification de document administratif est exclue. Le code des relations entre le public et l'administration nous précise même que ce document n'est pas communicable (article L311-5 CRPA 1°) « *Ne sont pas communicables : [...] les documents réalisés en exécution d'un contrat de prestation de services exécuté pour le compte d'une ou de plusieurs personnes déterminées* ». La notion de prestation de service employée ici est à comprendre au sens large : elle englobe également les contrats de recherche. Toutefois, cette interdiction peut être dépassée car l'accord du partenaire pour une diffusion du résultat en open data peut être obtenue.

Temps 4 : les exceptions

Une fois fixé sur la notion de document administratif, un examen « final » avant la mise en open data consiste à déterminer si l'on se trouve dans une hypothèse d'interdiction totale d'accès et de réutilisation. Nous aborderons ici quelques interdictions légales applicables aux productions d'un établissement de recherche :

- tout d'abord, comme examiné précédemment : les documents réalisés en exécution d'un contrat de prestation de services exécuté pour le compte d'une ou de plusieurs personnes privées déterminées ;
- ensuite ce qui porterait atteinte au secret de la défense nationale, à la sûreté de l'État, à la sécurité publique, à la sécurité des personnes, à la sécurité des systèmes d'information des administrations. Exemple : la divulgation de la localisation du stockage de certains virus.
- ce qui porterait atteinte aux secrets protégés par la loi, comme le secret médical, le secret des correspondances ...

Par ailleurs, certaines données ne sont communicables qu'à la personne concernée ou sous réserve de son accord. C'est le cas :

- des données qui portent atteinte au secret commercial et industriel. Cela va concerner par exemple des données transmises lors d'un partenariat de recherche, sous le sceau du secret ;
- des données personnelles. Ces données sont celles qui identifient directement ou indirectement une personne physique. Elles sont particulièrement présentes dans les bases de données dans le domaine de la recherche. Ainsi, une simple coordonnée GPS peut servir à identifier les propriétaires d'un terrain et à révéler leurs pratiques agricoles. Ce type de données est alors considéré comme identifiant indirectement une personne physique et peut ne pas être diffusable en open data ;
- des données sur lesquelles des tiers détiennent des droits de propriété intellectuelle.

Enfin, lors de la diffusion de données environnementales, il faudra veiller à ce que la diffusion des données n'apporte pas de nuisances aux éléments que l'on veut protéger. Exemple : diffusion des données de localisation de fleurs rares et afflux de curieux qui détruisent ces fleurs.

Les modalités de diffusion en open data

La diffusion des données en open data doit, on l'a vu, être gratuite et universelle.

Pour encadrer la **gratuité**, la réglementation oblige à choisir une licence de diffusion dans une liste imposée. Ainsi :

- pour les documents (rapports, etc.), la licence indiquée est la licence ouverte 2.0.
- pour les données disponibles sous forme de base de données, la licence recommandée est la licence ODBL 1.0 (open data base license).
- pour les logiciels, une liste de licences libres (avec ou sans copyleft) est disponible.

Ces recommandations sont issues d'un décret du 27 avril 2017. L'ensemble des licences à utiliser est consultable sur le site <http://www.data.gouv.fr/fr/licences>

Cas particulier : pour les logiciels, des vérifications supplémentaires doivent être effectuées avant la mise en accès libre, pour déterminer notamment si du code sous licence libre a été incorporé dans le logiciel.

Pour garantir un **accès universel** aux données, deux précautions sont à respecter :

- utiliser un format de fichier ouvert (c'est-à-dire interopérable avec tout système et ne nécessitant pas l'achat d'un logiciel particulier) ;
- mettre à disposition les données dans un entrepôt ouvert à tous, c'est-à-dire non limité à une communauté.

Le mot de la Direction des Affaires Juridiques de l'Inra

Cet article, qui décrit l'organisation de l'ouverture des données sous l'angle de son applicabilité aux productions de la recherche, illustre la complexité du sujet et de sa mise en œuvre. Le sujet est riche, aussi bien par sa nouveauté et ses enjeux, que par la réflexion qu'il nécessite pour appréhender ses aspects opérationnels, en particulier en termes d'aide à la décision.

Comment prendre position lorsque, à l'injonction d'ouverture répond une série d'exceptions ?

La Direction des Affaires Juridiques de l'Inra (« DAJ »), créée au mois d'avril 2018, propose ses ressources et son expertise pour contribuer à cet arbitrage, en pivot entre les différentes lectures opérationnelles de la réglementation.

Par sa vision globale du sujet, la DAJ veille ainsi à la sécurité juridique de l'ouverture des données conformément aux missions qui lui ont été confiées. Cet appui se veut concret et complémentaire aux initiatives existantes, dans l'objectif de rendre ce sujet lisible par tous.

Contacts :

Cécile Janet, Directrice des Affaires Juridiques (Cecile.janet@inra.fr)

Stéphanie Rennes, juriste en charge de l'Open Data à la Direction des Affaires Juridiques (Stephanie.rennes@inra.fr)

Référence bibliographique

Becard N, Castets-Renard C, Chassang G, Dantant M, Freyt-Caffin L, Gandon N, Martin C, Martelletti A, Mendoza-Caminade A, Morcrette N, Neirac C (2017) *Ouverture des données de la recherche. Guide d'analyse du cadre juridique en France.* doi :10.15454/1.481273124091092E12
<https://prodinra.inra.fr/record/382263>

Cet article est publié sous la licence Creative Commons (CC BY-SA).



<https://creativecommons.org/licenses/by-sa/4.0/>

Pour la citation et la reproduction de cet article, mentionner obligatoirement le titre de l'article, le nom de tous les auteurs, la mention de sa publication dans la revue « Le Cahier des Techniques de l'INRA », la date de sa publication et son URL).