

Ouverture des données publiques



Observations et propositions

Synthèse des propositions

Les données

- Imposer aux acteurs publics une obligation de mise à disposition des données publiques.
- S'assurer que les jeux de données mis à disposition soient complets, dans des formats exploitables, et que les données soient mises à jour et à disposition de manière régulière.
- Mener une réflexion sur l'intérêt de prévoir des délais différents de mise à disposition des données en fonction de la nature de celles-ci.
- Assurer l'interopérabilité des données.
- Encourager la définition de référentiels communs et s'assurer de leur application par tous.
- Supprimer les redevances pour les données publiques sauf si cette redevance emporte des conséquences économiques pour l'acteur public concerné.

Les acteurs publics

- Former les agents publics aux enjeux de l'open data et faire en sorte que la dimension open data soit intégrée systématiquement dans chaque action, programme ou politique publique.
- Créer un « référentiel communication » afin de réduire les délais de communication des documents administratifs.
- Élaborer des bonnes pratiques qui pourraient être diffusées par une structure de gouvernance partagée (comprenant à la fois des représentants d'émetteurs et des représentants d'utilisateurs des données). Ces bonnes pratiques devront prendre en compte la question de la protection de la vie privée.

La protection de la vie privée

- Accompagner, informer et sensibiliser les acteurs de l'open data (administrations et réutilisateurs) à cette question et aux solutions techniques d'anonymisation des données.
- Assurer aux citoyens une meilleure transparence sur la mise à disposition et la réutilisation des données publiques par l'État et les informer, les sensibiliser sur le fonctionnement de l'open data et de manière plus générale sur les usages et les outils numériques.
- Organiser des rencontres entre les acteurs afin de dégager des lignes directrices.
- Diffuser des guides recensant les bonnes pratiques, les outils et les solutions techniques mises à disposition.



Introduction

L'open data, ou ouverture des données publiques, poursuit plusieurs objectifs :

- moderniser l'action publique, favoriser la transparence comme l'efficacité de l'administration vis-à-vis du public (citoyens, entreprises...) et réduire les dépenses publiques ;
- stimuler la recherche et l'innovation : permettre à des entreprises de proposer de nouveaux services à valeur ajoutée.

Depuis la loi dite « CADA » de 1978¹, les citoyens peuvent accéder aux données produites par l'administration et les organismes publics.

En 2003, la directive 2003/98/CE du Parlement européen et du Conseil du 17 novembre 2003², a permis la réutilisation des documents et des informations publiques des organismes du secteur public. Récemment, la directive 2013/37/UE du Parlement européen et du Conseil du 26 juin 2013 modifiant la directive 2003/98/CE concernant la réutilisation des informations du secteur public propose un cadre juridique harmonisé à l'échelle de l'Europe. Elle prévoit notamment une extension du champ d'application aux bibliothèques, musées et archives avec toutefois un certain nombre de limitations ; la mise à disposition des informations publiques dans un format ouvert et lisible par machine ; des redevances limitées aux coûts marginaux de reproduction, de mise à disposition et de diffusion. Cette directive devra être appliquée par les États membres à partir du 18 juillet 2015.

De nombreuses initiatives ont contribué au développement de l'open data en France (création de la mission Etalab en 2011, feuille de route du Gouvernement en matière d'ouverture et de partage des données publiques »...).

Pour les entreprises, l'ouverture des données dans des conditions optimum d'exploitation est impérative pour encourager l'innovation et la création de valeur par les acteurs économiques. Cependant, dans les faits, les expériences menées par les entreprises intervenant dans l'exploitation des données publiques, conduisent à identifier un certain nombre de freins et donc de pistes d'amélioration des processus en vigueur.

À noter, la notion de confiance doit être posée en filigrane de tous les éléments de la réflexion. Celle-ci conditionne autant les comportements des administrations, des citoyens, que ceux des entreprises.

À ce stade, les « success stories » de l'open data à la française sont rares et démontrent bien que des aménagements sont encore nécessaires en la matière.

¹ Loi n° 78-753 du 17 juillet 1978 portant diverses mesures d'amélioration des relations entre l'administration et le public et diverses dispositions d'ordre administratif, social et fiscal.

² Directive transposée par l'ordonnance n° 2005-650 du 6 juin 2005 relative à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques et le décret n° 2005-1755 du 30 décembre 2005 relatif à la liberté d'accès aux documents administratifs et à la réutilisation des informations publiques, pris pour l'application de la loi n° 78-753 du 17 juillet 1978.



1. Sur les données

Les données publiques sont définies comme « les documents existants détenus par des organismes du secteur public des États membres » (article 1.1 de la directive du 17 novembre 2003 concernant la réutilisation des informations du secteur public et non modifié par la directive du 26 juin 2013).

On entend par « organismes du secteur public », l'État, les collectivités territoriales, les organismes de droit public et les associations formées par une ou plusieurs de ces collectivités ou un ou plusieurs de ces organismes de droit public (article 2.1 de la directive du 17 novembre 2003 concernant la réutilisation des informations du secteur public et non modifié par la directive du 26 juin 2013).

Ces données peuvent être réutilisées par toute personne à d'autres fins que celles initialement prévues dans le cadre de la mission de service public. Toutefois, un certain nombre de données sont exclues de toute réutilisation comme par exemple les documents susceptibles de porter atteinte au secret de la défense nationale, à la sûreté de l'État, sur lesquels des tiers détiennent des droits de propriété intellectuelle, lorsqu'ils comportent des données à caractère personnel...

Outre ces exceptions, la mise en pratique de cette réutilisation se heurte à des freins rendant la réexploitation des données difficile voire impossible. En effet, force est de constater que les données sont parfois inaccessibles, incomplètes, non mises à jour, mises à disposition tardivement, payantes, techniquement inexploitable...

a- Des données inaccessibles

Il existe des données parfaitement anonymes qui ne sont pas accessibles.

Par exemple, l'arrêté relatif au Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM) définit qui a le droit d'accéder aux données de remboursement de l'assurance maladie obligatoire et à quelles données (champ restreint).

b- Des données incomplètes

En dépit de l'effort des acteurs publics consistant à mettre à disposition un nombre important de données, il apparaît toutefois que cette mise à disposition nécessite d'être optimisée. En effet, des jeux de données sont encore trop souvent incomplets pour permettre leur exploitation par les entreprises. Ces mises à disposition partielles de données peuvent résulter d'une difficulté technique, d'une méconnaissance par les acteurs publics de leur intérêt ou encore d'un doute concernant la possibilité de les rendre publiques.

Ainsi, par exemple, la base des médicaments fournie en open data par l'agence nationale de sécurité du médicament et des produits de santé (ANSM) ne comprend pas les notices des médicaments qui elles, sont publiées sur un autre site et disponibles en simple consultation et non téléchargeable.



Aussi, est-il impératif que les jeux de données mis à la disposition soient complets, c'est-à-dire qu'ils comprennent tous les éléments dont dispose l'émetteur dans la limite des données publiques pouvant être mises à disposition.

c- Des données non mises à jour ou mises à jour tardivement

La question de la mise à jour des données et de la régularité de leur mise à disposition est capitale. En effet, de ces éléments dépendent la pertinence, la qualité et la fiabilité voire l'existence même des applications qui découlent de la réutilisation de ces données.

Par exemple, les taux des impôts locaux pour 2012 ont été mis à disposition sur le site impots.gouv.fr très tardivement (mars 2014).

d- Des délais de mise à disposition des données trop longs ou trop aléatoires

Dans le droit fil de la logique de mise à jour régulière, la durée entre la compilation des données et leur mise à disposition constitue également un frein à l'exploitation économique de ces données. Des délais de publication variables d'une année à l'autre rendent très hypothétique et peu viable un modèle reposant sur ces données.

À titre d'exemple, les dernières données accessibles à ce jour sur le site impots.gouv.fr, concernant les impôts sur le revenu, sont celles des impôts sur le revenu 2011 calculés sur les revenus 2010.

Une réflexion pourra être menée sur l'intérêt de prévoir des délais de mise à disposition différents en fonction de la nature des données.

e- Une trop faible interopérabilité des données entre elles

Pour être réutilisées, les données doivent être structurées (informations organisées et classées en vue de faciliter leur lecture et leur traitement). En outre, les recherches et classifications doivent être facilitées pour une réutilisation et une exploitation des données. La question de l'harmonisation des structures des données entre les administrations est donc importante.

Or, il n'existe pas un référentiel commun partagé par les acteurs publics ou lorsqu'il existe des référentiels, ceux-ci ne sont pas systématiquement utilisés.

Par exemple, le code INSEE pour les villes n'est pas systématiquement utilisé dans les jeux de données, empêchant en pratique tout croisement.

Il est nécessaire d'encourager la définition de référentiels communs et de s'assurer de leur application par tous (code INSEE pour les villes, référentiel GPS identique pour la localisation...).

f- Des données payantes

Les frais liés à la délivrance d'une copie d'un document administratif sont supportés par le demandeur, ces frais ne pouvant excéder le coût de la reproduction (article 4-b) de la loi du 17 juillet 1978 modifiée et article 35 du décret n° 2005-1755 du 30 décembre 2005). Le droit de réutilisation d'informations publiques, quant à lui, peut donner lieu au versement de redevances.

Le Gouvernement a réaffirmé son attachement à la gratuité de la réutilisation des données publiques à l'occasion du Comité interministériel pour la modernisation de l'action publique (CIMAP) du 18 décembre 2013 ainsi que dans la « Stratégie gouvernementale en matière d'ouverture et de partage des données publiques » publiée le 28 février 2013. Sur la base du rapport remis par M. Trojette relatif au modèle économique des redevances de réutilisation des données publiques, le Gouvernement a décidé de supprimer plusieurs redevances et de préciser sa doctrine en matière d'exceptions au principe de gratuité.

La directive 2013/37/UE du Parlement européen et du Conseil du 26 juin 2013 prévoit des redevances limitées aux coûts marginaux de reproduction, de mise à disposition et de diffusion (articles 6 et 7).

En dépit de ces efforts qu'il convient de saluer, des jeux entiers de données restent encore soumis à redevance, comme par exemple, la base des permis de construire ou les annonces légales. Le paiement d'une redevance reste un frein car il implique un risque supplémentaire en termes de rentabilité supposée de l'investissement pour les acteurs souhaitant développer des usages open data. En effet, l'achat de données, quand ni leur régularité, ni leur format, ni leur interopérabilité ne sont garantis, contribue à restreindre le nombre d'entreprises qui souhaiteraient se positionner sur le marché de l'open data.

En outre, le recouvrement de cette redevance peut parfois générer pour les acteurs publics, des coûts supérieurs à ladite redevance.

Aux États-Unis, les politiques publiques de mise à disposition des données font le pari de la gratuité pour impulser le développement d'affaires à partir des données publiques.

Au vu de ces éléments, les données publiques ne devraient pas, par principe, être soumises à redevance sauf si cette redevance emporte des conséquences économiques pour l'acteur public concerné.

2. Sur les acteurs publics de l'open data

a- Une trop faible sensibilisation des enjeux

Certains émetteurs potentiels, par manque d'informations ou par manque de moyens (humains et temps) sont freinés dans la pratique systématique de la mise à disposition de leurs données.

Si la présence d'un référent « open data » au sein de chaque administration est *a priori* acquise (cf. décret de 2005 de la loi de juillet 1978), **il n'en demeure pas moins qu'une formation plus large des agents publics aux enjeux de l'open data serait un plus incontestable et un préalable à toute volonté de systématiser cette pratique au sein des administrations.**

Il s'agit là, de travailler sur la levée des blocages culturels qui freinent l'administration dans sa dynamique d'open data. Le rapport d'information de la mission d'information sur l'accès aux



documents administratifs et aux données publiques du 12 juin 2014³ estime que les administrations sont « souvent peu diligentes, frileuses, voire de mauvaise volonté ». Il note « une inertie persistante d'un certain nombre de services et une incontestable frilosité, qui se traduit par une interprétation extensive de la portée des secrets et intérêts protégés par la loi [...] ».

Par ailleurs, il conviendrait de **s'assurer que la dimension open data soit intégrée systématiquement dans chaque action, programme ou politique publique.**

b- Une réduction des délais de communication des documents administratifs

Il pourrait être envisagé, comme le recommande le rapport de la mission commune d'information sur l'accès aux documents administratifs et aux données publiques du Sénat intitulé « Refonder le droit à l'information publique à l'heure du numérique : un enjeu citoyen, une opportunité stratégique » du 12 juin 2014 (recommandation principale n°1 – pages 147 et 148) , de **créer « un référé communication »** devant le juge administratif sur saisine de la CADA dès la réception par celle-ci de la demande d'avis « lorsque la question de la communicabilité du document auquel l'accès est demandé a déjà été tranchée par la CADA ou la jurisprudence ou si le document figure sur une liste fixée par arrêté après avis de la CADA ».

c- Un levier de modernisation

Une approche open data favorise une collaboration nouvelle entre les acteurs publics ou encore entre leurs services internes comme en attestent certaines expériences étrangères (cf. Royaume-Uni). Elle peut en outre créer un processus unifié de production de données et fournir des informations nouvelles facilitant notamment la gestion des acteurs contributeurs. Dans ces conditions, l'open data crée non seulement de la valeur pour les citoyens mais **constitue également un levier d'efficience des administrations publiques.**

Par ailleurs, l'open data peut permettre de fédérer les différents acteurs des administrations publiques.

Ces collaborations vont faciliter le partage, la diffusion et la réutilisation des informations produites entre les acteurs publics ou/et entre leurs services et les aider à être davantage performants dans leur mission de service public.

Le décloisonnement des données issues de diverses applications internes donne en effet aux acteurs publics **l'opportunité d'agréger leurs données, d'en faire ressortir des informations pertinentes et d'améliorer son fonctionnement et ses actions.**

d- Une gouvernance structurée

Le constat de disparité des formats, de nature ou encore de décomposition des données d'un organisme public à l'autre devrait conduire à **l'élaboration de bonnes pratiques qui pourraient**

³ Rapport d'information de la mission commune d'information sur l'accès aux documents administratifs et aux données publiques du Sénat intitulé « Refonder le droit à l'information publique à l'heure du numérique : un enjeu citoyen, une opportunité stratégique » du 12 juin 2014 :

http://www.senat.fr/commission/missions/acces_aux_documents_administratifs_et_aux_donnees_publicques/index.html

être diffusées par une structure de gouvernance partagée (comprenant à la fois des représentants d'émetteurs et des représentants d'utilisateurs des données). Dans un contexte européen voire mondialisé, il apparait alors inévitable de **suggérer la définition de référentiels communs** qui pourraient non seulement permettre l'interopérabilité des données entre elles, mais également leur comparaison, leur mutualisation et la création de nouveaux services de portée internationale.

e- Un principe de mise à disposition obligatoire des données publiques

Au vu des éléments ci-dessus, **il pourrait être envisagé d'imposer aux acteurs publics une obligation de mise à disposition des données publiques.**

3. Sur la protection de la vie privée

La loi du 17 juillet 1978 modifiée (dite « CADA ») ainsi que le décret du 30 décembre 2005 prévoient des dispositions concernant la mise à disposition et la réutilisation d'informations publiques comportant des données personnelles.

La réutilisation des informations publiques comportant des données à caractère personnel n'est possible que si le consentement de la personne concernée a été recueilli ou si les données sont anonymisées ou si une disposition législative ou réglementaire permet la réutilisation (conditions alternatives). En outre, l'article 13 précise que la réutilisation d'informations publiques comportant des données à caractère personnel est subordonnée au respect des dispositions de la loi n° 78-17 du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés.

a- Un cadre juridique protecteur

Un cadre juridique existe (loi dite « CADA » et loi dite « Informatique et libertés ») et est protecteur en matière de données personnelles.

La CADA ainsi que le juge administratif ont progressivement circonscrit dans des situations données des éléments à prendre en compte au titre de la vie privée (par exemple, avis de la CADA 20121581 - séance du 19/04/2012⁴, arrêt de la Cour administrative d'appel de Lyon 3ème chambre du 4 juillet 2012⁵...). Les approches de la CADA et de la CNIL doivent nécessairement s'imbriquer dans un ensemble cohérent et transparent pour les acteurs.

Les logiques qui sous-tendent la mise à disposition en vue de la réutilisation des données publiques et la protection des données personnelles sont différentes : les données publiques sont mises à disposition et peuvent être réutilisées par des tiers pour des finalités qui ne sont pas connues au moment de leur mise à disposition. La licéité et le caractère loyal du traitement de données à caractère personnel sont appréciés quant à eux en fonction des grands principes de la protection des données personnelles : légitimité, consentement, proportionnalité, finalité. Les

⁴ <http://www.cada.fr/avis-20121581,20121581.html>

⁵ <http://www.legifrance.gouv.fr/affichJuriAdmin.do?oldAction=rechJuriAdmin&idTexte=CETATEXT000026206985&fastReqId=564239878&fastPos=1>

acteurs utilisateurs des données publiques en open data doivent pouvoir être à même d'évaluer la nature des obligations qui pèseraient sur eux lors des usages de données notamment au regard des droits des personnes. En raison de l'interprétation extensive de la notion de donnée à caractère personnel (de plus en plus de données à caractère technique sont incluses dans le champ des données personnelles), il peut en découler des obligations sanctionnables que les utilisateurs doivent être en mesure d'appréhender.

b- Des confusions à éviter

Des distinctions doivent être effectuées entre les notions de big data, d'open data et de données à caractère personnel.

Le **big data** peut se définir comme une utilisation massive de données structurées ou non structurées (capteurs utilisés pour collecter les informations climatiques, messages sur les sites de médias sociaux, images numériques et vidéos publiées en ligne, signaux GPS de téléphones mobiles...). Les données ne sont pas nécessairement des données publiques ou des données personnelles.

Il est possible avec la technique du big data de ré-identifier des personnes à partir de données publiques recoupées/croisées avec d'autres données : en réalité, il ne s'agit pas d'un problème lié à la mise à disposition des données publiques (le croisement des données n'est pas lié à l'open data) mais d'un problème lié à la manière d'anonymiser les données publiques mises à disposition. Par exemple, l'INSEE met à disposition les données du montant d'impôt par zone géographique. La superposition de cartes géographiques (Google Maps...) a permis de se rendre compte que certaines zones géographiques ne comportent qu'un seul résident. Ici, on voit le souci lié à la libéralisation d'un jeu de donnée où pour une entrée géographique il n'y avait qu'un résident fiscal. Autre exemple : décision de justice comportant le prénom et des indications sur la fonction (actuel président de la République). Là aussi, le problème se situe dans la manière d'anonymiser les données. Il est absolument nécessaire de disposer de systèmes d'anonymisation particulièrement développés et sécurisés.

L'**open data** consiste en la mise à disposition des données publiques qui sont définies comme « les documents existants détenus par des organismes du secteur public des États membres » (article 1.1 de la directive du 17 novembre 2003 concernant la réutilisation des informations du secteur public et non modifié par la directive du 26 juin 2013).

On entend par « organismes du secteur public », l'État, les collectivités territoriales, les organismes de droit public et les associations formées par une ou plusieurs de ces collectivités ou un ou plusieurs de ces organismes de droit public (article 2.1 de la directive du 17 novembre 2003 concernant la réutilisation des informations du secteur public et non modifié par la directive du 26 juin 2013).

Par principe, les données publiques ne sont pas des données à caractère personnel (anonymisation) sauf exceptions (arrêtés de nomination, annuaire de « Service-Public », des données – liste des personnes auditionnées par exemple - contenues dans des rapports publics, des comptes rendus de réunion, des débats...). Donc très peu de données publiques comportent des données à caractère personnel.



Une **donnée à caractère personnel** est toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement. Dans le cadre de l'open data, des données à caractère personnel peuvent concerner des entrepreneurs individuels ou encore des agents publics. Il convient de respecter le principe de proportionnalité : par exemple, le nom d'un entrepreneur individuel ou d'un agent public constitue à la fois une donnée personnelle et une donnée professionnelle, ce qui rend légitime sa diffusion dans les registres publics.

Il est nécessaire de ne pas confondre l'open data et la protection des données personnelles en ce que cela pourrait constituer un frein à la mise à disposition des données et à leur réutilisation. Il importe d'avoir une approche pragmatique, cohérente et de coopération entre les acteurs et les autorités. Au-delà, les vocations et les usages de l'open data n'ayant pas de frontières, l'approche cohérente amorcée notamment par le G29 pour identifier les points de vigilance liés à la mise à disposition et la réutilisation des données est à encourager (cf. avis 06/2013 du G29 du 5 juin 2013).

Il conviendrait :

- **d'accompagner, informer et sensibiliser les acteurs de l'open data (administrations et réutilisateurs) à cette question et aux solutions techniques d'anonymisation des données⁶.**
- **d'assurer aux citoyens une meilleure transparence sur la mise à disposition et la réutilisation des données publiques par l'État et les informer, les sensibiliser sur le fonctionnement de l'open data** et de manière plus générale sur les usages et les outils numériques. **Des rencontres entre les acteurs pourraient être organisées⁷** afin de permettre de dégager des lignes directrices et donner l'opportunité de débats fructueux et constructifs entre autorités compétentes et acteurs professionnels en particulier pour les données dont le caractère personnel n'est parfois pas aisé à percevoir. À l'instar des guides de sécurité publiés en 2012 par la CNIL, **des guides recensant les bonnes pratiques, les outils et les solutions techniques mises à disposition peuvent être une voie encourageante pour l'innovation.** Il faut continuer d'encourager ces démarches plutôt que de se précipiter à légiférer.

⁶ Voir l'étude annuelle 2014 du Conseil d'État sur le numérique et les droits fondamentaux de septembre 2014 : <http://www.conseil-etat.fr/Decisions-Avis-Publications/Etudes-Publications/Rapports-Etudes/Etude-annuelle-2014-Le-numerique-et-les-droits-fondamentaux> et le rapport de la Commission Open data en santé du 9 juillet 2014 : <http://www.drees.sante.gouv.fr/rapport-de-la-commission-open-data-en-sante,11323.html>.

⁷ Comme par exemple, la journée sur l'open data organisée à l'initiative de la CNIL en juillet 2013 et la consultation récente de janvier 2014.