RÉPUBLIQUE FRANÇAISE

LA PROTECTION DES DONNÉES PERSONNELLES DANS L'*OPEN DATA* : UNE EXIGENCE ET UNE OPPORTUNITÉ

COMMISSION DES LOIS
Rapport d'information de MM. Gaëtan Gorce et François Pillet

■ La France s'est résolument engagée sur la voie de l'ouverture et du partage des données publiques, plus connue sous le nom d'*open data*.
Deux idées animent cette politique. Comptables de leur gestion auprès des citoyens, les administrations leur ouvrent leurs fichiers. Elles leur donnent ainsi le moyen de mieux les contrôler. Par ailleurs, à l'ère du numérique, où l'information est source de richesse, elles leur offrent l'opportunité d'exploiter le formidable gisement que constituent ces données.
En créant une mission d'information chargée d'étudier l'*open data* et la protection de la vie privée de nos concitoyens, la commission des lois a souhaité poursuivre sa réflexion sur les nouveaux usages numériques et la façon dont ils peuvent se concilier avec les principes fondamentaux que le législateur a posés dès la fin des années 1970.
L'*open data* soulève à cet égard une question spécifique : en principe, il exclut toute diffusion de données à caractère personnel, mais bien souvent, les données détenues par les administrations ont été élaborées à partir d'informations individuelles, qui peuvent être retrouvées grâce aux formidables capacités de traitement que permet l'informatique moderne. L'impératif de protection de la vie privée est-il en mesure de toujours prévaloir ? Comment s'en assurer ?
■ À l'issue de leurs travaux, les rapporteurs de la mission d'information, les sénateurs Gaëtan Gorce (Soc. – Nièvre) et François Pillet (ratt. UMP – Cher) jugent aujourd'hui nécessaire de faire d'une exigence fondamentale –la protection de la vie privée de nos concitoyens– une opportunité pour donner une nouvelle impulsion au déploiement de l'*open data*.

L'*open data* : en dépit d'un cadre réglementaire protecteur des données personnelles, une double faille à corriger

Mouvement récent, l'*open data* est aujourd'hui un des axes importants de la modernisation de l'action publique de nombreux pays d'Europe et des États-Unis.
Il recouvre deux principes : le premier est la mise en ligne, par les administrations, des données qu'elles détiennent ; le second, la libre réutilisation par les citoyens ou les entreprises des données ainsi publiées. Il est inspiré par deux objectifs : garantir la transparence de l'action publique en permettant à chacun de consulter les données relatives à l'action de l'administration et les informations sur lesquelles elle fonde ses décisions ; offrir à tous la possibilité d'exploiter ces données et d'en tirer un profit pour soi ou le bien commun.

Cet *open data* prend aujourd'hui la forme, en France, d'une ambition portée par plusieurs gouvernements successifs, d'une pratique qui se diffuse progressivement au sein des administrations, et d'une structure chargée de fédérer et d'inciter les initiatives, la mission *Etalab*.

■ Un cadre juridique en principe protecteur des données personnelles…

L'*open data* repose, en France, sur les deux lois pionnières de 1978, la loi « *Informatique et libertés* » du 6 janvier et la loi « *CADA* » du 17 juillet sur l'accès aux documents administratifs, modifiée au cours des années 2000, pour traiter spécifiquement, dans le respect des principes antérieurs, de la question de la réutilisation des informations publiées par les administrations.

Ce cadre juridique apporte une triple garantie à la protection des données personnelles. La première tient à l'interdiction de principe qu'une donnée personnelle fasse l'objet d'une mise en ligne par l'administration et d'une réutilisation par un tiers. Ce principe connaît trois exceptions : le consentement de l'intéressé à cette diffusion, l'existence d'une obligation légale de publication, ou –exception la plus générale et la plus commode en matière d'*open data*–l'anonymisation des données publiées. La seconde garantie correspond à la soumission de toute réutilisation de données personnelles aux exigences de la loi « Informatique et libertés ». Enfin, la dernière garantie est celle de la sanction qui pourrait frapper le non-respect des dispositions précédentes : engagement de la responsabilité de l'État, voire condamnation pénale pour diffusion de données personnelles par négligence.

Les rapporteurs constatent à cet égard, qu'à l'exception de rares accidents, l'*open data* pratiqué en France n'a pas, jusqu'à présent, posé de graves problèmes pour la protection de la vie privée des citoyens.

■ ... Toutefois fragilisé par une double faille : un risque de ré-identification avéré et un défaut de pilotage

La situation n'est toutefois pas satisfaisante et les rapporteurs ont identifié deux failles, qui fragilisent le dispositif actuel.

En effet, tout repose sur la solidité de l'anonymisation par l'administration des jeux de données qu'elle publie. Or, du fait des capacités de croisement des informations qu'autorise l'informatique moderne, le risque d'une ré-identification des données publiées existe. Il se trouve même aggravé par la profusion de jeux de données mis en ligne par l'administration comme par les personnes privées elles-mêmes.

La façon dont l'État et les collectivités territoriales conduisent l'ouverture de leurs données devrait pouvoir dissiper les inquiétudes. Il n'en est rien : le sentiment prédomine d'un défaut de pilotage ou d'accompagnement qui laisse parfois les administrations démunies face à une tâche nouvelle qu'elles ne maîtrisent pas toujours. C'est ainsi qu'en 2013, l'identité et l'imposition de certains contribuables ont pu être retrouvées dans une base pourtant anonymisée, parce que le procédé utilisé, qui consistait à agréger toutes les impositions des contribuables habitant la même zone géographique de 200 m sur 200 m (technique du carroyage), était appliqué à des zones très peu peuplées.

Poursuivre le développement de l'*open data* en l'assortissant de garanties plus solides pour la protection des données personnelles

Ces failles ne remettent pas en cause la pertinence de l'ouverture des données publiques, mais la façon dont elle est conduite.

Loin de trouver là des raisons de freiner un mouvement dont l'utilité sociale est acquise, la mission d'information y a plutôt vu l'opportunité de donner un nouvel élan à l'ouverture et au partage des données publiques, en définissant une doctrine et une méthode qui garantissent la meilleure protection des données personnelles possible. Car, une fois cette protection assurée, aucun obstacle au déploiement de l'*open data* n'est plus légitime.

■ Une stratégie : faire de l'*open data* la règle

Les rapporteurs recommandent donc de poser le principe d'une obligation de mise en ligne des données détenues par les administrations, à moins que le coût en soit trop important ou que les risques pour la vie privée ne puissent être levés par une anonymisation efficace.

Ceci suppose de prévoir une période transitoire pendant laquelle les administrations achèveront le recensement des jeux de données qu'elles détiennent, indiqueront ceux qui ne pourront faire l'objet d'une publication pour une des raisons précitées et élaboreront, pour les autres, un calendrier de mise en ligne.

■ Mettre en œuvre une doctrine de protection des données personnelles

Les administrations doivent s'investir dans cette voie et mettre en œuvre une doctrine de protection des données personnelles, en anticipant et évaluant les risques, en y adaptant les formats de diffusion des données et en exerçant une veille vigilante.

Il s'agit notamment de concevoir, dès l'origine, les bases de données dans la perspective de leur mise en ligne, afin d'en faciliter l'anonymisation, de réaliser, pour chaque base de données, une étude d'impact sur le risque de ré-identification ou les difficultés d'anonymisation, de soumettre les jeux de données publiés à une surveillance régulière, et, si une fuite se produit, de définir à l'avance une stratégie de rapatriement des données compromises.

■ Adapter la gouvernance de l'*open data* à la protection des données personnelles

Les administrations doivent être secondées dans cette tâche par l'État, qui devra veiller à leur apporter, par une structure dédiée, au sein d'*Etalab*, une assistance technique, organisationnelle et juridique, en particulier s'agissant de l'évaluation des risques de ré-identification et de la mise en œuvre des procédés d'anonymisation.

L'anonymisation étant d'ailleurs la clé d'un *open data* respectueux des données personnelles, il est absolument nécessaire d'en garantir le financement. Il peut à cet égard paraître plus raisonnable de prévoir le paiement d'une redevance par les réutilisateurs que de renoncer, en raison d'un problème de financement des mesures d'anonymisation, à publier un jeu de données. D'autres voies de financement peuvent aussi être explorées, notamment celles du financement coopératif.

*
* *

Trouver dans la protection des données personnelles le levier pour porter plus haut l'exigence de transparence de l'action publique : le pari est audacieux, mais il est conforme à l'ambition que notre pays a démontrée depuis les lois pionnières de 1978.

Les principales recommandations de la mission d'information

Accélérer le déploiement d'un *open data* respectueux de la protection des données personnelles

1. Poser le principe que l'administration est tenue de mettre en ligne progressivement, en les anonymisant si nécessaire, toutes les bases de données qu'elle détient et qui seraient susceptibles d'être communiquées à un citoyen s'il en fait la demande ou qui font l'objet d'une diffusion publique sur un autre support. L'administration ne pourrait s'y opposer qu'en raison des coûts déraisonnables de gestion que cette mise en ligne imposerait (notamment les coûts d'anonymisation éventuelle), ou du risque avéré, qu'en dépit des précautions prises, des informations personnelles puissent être ré-identifiées.

Mettre en œuvre une doctrine de protection des données personnelles en matière d'*open data*

2. **Anticiper et évaluer** : prévoir, dès la conception de la base, dans la perspective de sa possible ouverture :

- les modalités de son anonymisation éventuelle ;
- le cas échéant, le marquage des jeux de données afin d'être en mesure de suivre les réutilisations éventuelles et dénoncer les mésusages.

Procéder, préalablement à tout examen de l'opportunité d'ouvrir une base de données, ainsi, le cas échéant, qu'à intervalles réguliers, à une analyse du risque de ré-identification et des conséquences possibles d'une telle ré-identification.

3. Adapter la diffusion en fonction du risque : en cas de risque avéré sur les données personnelles, impossible à éliminer par des procédés d'anonymisation, refuser l'ouverture des données ou, si le bénéfice social attendu de cette ouverture est jugé trop important, procéder à une ouverture restreinte de cette base.

Concevoir à cette fin un *continuum* de solutions d'accès aux données, allant de l'*open data* jusqu'aux modes d'accès les plus sélectifs.

4. Assurer une veille sur la diffusion et les réutilisations des données publiques : faciliter notamment les procédures par lesquelles un réutilisateur peut alerter l'administration compétente

Prévoir que l'administration définisse une stratégie de rapatriement ou de suppression des jeux de données compromis, afin de remédier rapidement à la diffusion accidentelle d'informations personnelles.

5. Renforcer la protection offerte par la licence de réutilisation : exclure expressément les données personnelles du champ d'application de la licence ouverte utilisée par les administrations pour la réutilisation des données publiques.

Interdire expressément dans le contrat de licence toute réutilisation abusive qui aboutirait à lever l'anonymisation des données.

Intégrer au contrat de licence, une clause de suspension légitime du droit de réutilisation ainsi que de suppression ou de rapatriement des jeux de données compromis lorsqu'un risque de ré-identification est apparu.

Adapter la gouvernance de l'*open data* aux exigences de la protection des données personnelles

6. Mettre en place, auprès de la mission *Etalab*, une structure dédiée à la protection des données à caractère personnel et chargée d'assister les administrations :

- dans l'élaboration de l'étude d'impact préalable à la mise à disposition des données ;
- dans l'anonymisation éventuelle de la base ;
- dans la mise en place d'un mode d'accès restreint.

7. Garantir le financement par l'État des mesures d'anonymisation des données personnelles contenues dans des jeux de données publiques.

Ne pas renoncer par principe au prélèvement d'une redevance en présence de coûts d'anonymisation élevés.

Encourager le financement coopératif de l'anonymisation.

Commission des lois

http://www.senat.fr/commission/loi/index.html

Téléphone : 01 42 34 23 37 – Télécopie : 01 42 34 31 47

Rapporteur
Gaëtan Gorce
Sénateur (socialiste) de la Nièvre

Rapporteur
François Pillet
Sénateur (ratt. UMP) du Cher