



Regards Citoyens  
c/o Laurent GUERBY  
10 chemin Tricou  
31200 TOULOUSE

À Paris, le 17 février 2014

Monsieur le Sénateur Gaëtan Gorce,  
15 rue de Vaugirard  
75291 Paris Cedex 06

Monsieur le Sénateur,

Vous nous avez conviés à une audition au Sénat le 29 janvier 2014. Ayant exprimé par le passé des positions assez surprenantes en vous opposant franchement à l'Open Data, nous nous réjouissons que vous nous offriez l'opportunité de pouvoir vous expliquer les motifs de notre profond désaccord avec les positions que vous aviez exprimées et de porter à votre connaissance des éléments qui auraient pu vous faire défaut. Nous avons en effet le sentiment d'observer une forme de confusion entre deux notions assez nouvelles : l'Open Data - cette démarche démocratique mise en œuvre par un nombre croissant d'institutions pour assurer la libre réutilisation des données publiques, ne contenant donc pas de données personnelles - et le Big Data - cette technologie notamment très utilisée dans le traitement de données personnelles.

Nous avons de bon cœur pris une demi-journée de congés pour pouvoir libérer du temps à nos activités bénévoles. C'est donc profondément déçus que nous nous sommes retrouvés face à votre co-rapporteur M. Pillet sans qu'aucune explication ne nous soit formulée pour excuser votre absence.

Cette audition n'ayant par ailleurs pas été enregistrée au contraire de celle organisée par vos collègues de la mission d'information dédiée aux documents administratifs et aux données publiques, nous nous permettons de vous adresser les quelques notes que nous avons préparées en vue de notre rencontre.

En espérant que ces éléments puissent alimenter vos réflexions personnelles, nous vous prions d'agréer, Monsieur le Sénateur, l'expression de nos meilleurs sentiments.

L'équipe de Regards Citoyens

***Mission d'information sénatoriale « Open Data et données personnelles »***  
***Notes préparatoires à l'audition de Regards Citoyens***

Regards Citoyens - 28 janvier 2014



Des questions nous ayant été soumises en amont de l'audition, nous l'avons préparée en suivant ce canevas. Les questions fournies en amont de notre audition apparaissent en gras dans la suite du document.

En guise d'introduction, nous tenons à nous émouvoir du titre de la mission d'information qui vise à créer une confusion entre Open Data et données personnelles. La définition même de l'Open Data est très claire : c'est la libre réutilisation des données publiques non personnelles (des noms peuvent être rendus publics sous des conditions très strictes, dans ce cas, on parle de données nominatives). Si nous comprenons que cet amalgame soit entretenu par certains lobbyistes qui défendent les intérêts d'entreprises qui s'enrichissent en collectant de très nombreuses données personnelles, le fait que des institutions publiques participent à cet amalgame est très inquiétant.

### **Les principes retenus par la législation nationale ou européenne en vigueur vous semblent-ils suffisants pour garantir la protection de la vie privée dans le cadre de la mise à disposition des données publiques ?**

Une des chances de la France est qu'elle a une des lois les plus protectrices en matière de divulgation des données personnelles. Sauf exception prévue par le législateur, les données publiques mises à disposition pour réutilisation ne peuvent contenir des données à caractère personnel, c'est à dire des données permettant de remonter à une personne (et pas seulement les données d'identification d'une personne). De plus la loi CADA, encadrant la réutilisation de données publiques, prévoit un strict respect des dispositions liées au secret statistique : toute divulgation de données basées sur l'agrégation de données personnelles anonymisées doit passer le contrôle d'une commission qui veille à ce que la désanonymisation ne soit pas possible. Le dispositif législatif français assure donc que les données librement réutilisables ne contiennent pas de données à caractère personnel.

### **Quels types de données publiques, construites à partir de données personnelles vous semblent particulièrement sensibles de ce point de vue ?**

La définition même de l'Open Data fait une claire distinction entre données publiques librement réutilisables et données personnelles (qui doivent rester maîtrisées par les seuls créateurs de ces données). L'Open Data n'est donc pas un danger.

En revanche, il existe des problèmes de vente de données publiques à des tiers qui ne répondent pas aux critères attendus par la loi :

- les données liées à la doctrine ou la jurisprudence dont l'anonymisation est confiée aux marchands de données juridiques (LexisNexis, Lefebvre, Lamy...) et dont la diffusion de données non anonymisées n'est absolument pas maîtrisée par les acteurs publics ;
- les données du système d'immatriculation des véhicules (cartes grises) ont été vendues en toute illégalité pendant des années et la légalisation de leur revente n'a pas réglé le problème de divulgation des données personnelles ;
- les données carroyées des ménages ont été vendues pendant des années sans que le niveau d'anonymisation soit satisfaisant (donc en méconnaissance du secret statistique). C'est justement à l'occasion des travaux préparatoires au passage en Open Data que ces gros problèmes ont pu être identifiés et finalement corrigés.

Avec ce dernier exemple, on voit que l'Open Data est source d'un meilleur respect des données personnelles en permettant de détecter le non respect des règles en vigueur en termes d'anonymisation et de respect des données personnelles.

### **Comment est assuré le respect des droits des personnes dans la diffusion et réutilisation plus ou moins directes de leurs données (consentement, opposition, rectification...) ?**

La règle est que la diffusion de données publiques implique l'absence d'information à caractère personnel. Il faut faire la différence entre données personnelles et données nominatives. La loi CADA prévoit, pour des raisons d'intérêt général, la diffusion de certaines données nominatives. C'est par exemple le cas des noms des élus afin de permettre à tous les citoyens de suivre leurs activités électives. Dans ce cas, le consentement ou l'opposition ne sont pas à obtenir car il est organisé par le droit. Le droit de rectification reste opposable en cas d'erreur.

### **Les législations CADA et CNIL vous semblent-elle s'articuler correctement sur la question de la mise à disposition d'informations publiques et de la protection des données personnelles ? Même question s'agissant des régimes spéciaux de publicité légale ?**

La hiérarchie des normes offre un cadre d'analyse plus large à la CADA par rapport à la CNIL. En assurant des

régimes d'exception clairs à la loi CNIL, la loi CADA assure que la transparence de la prise de décision publique puisse être garantie dans le respect de la vie privée des acteurs impliqués : ce n'est pas parce qu'une disposition prévoit, pour des raisons d'intérêt général, la publication de noms que la vie privée de ces personnes est atteinte. C'est en tout cas ce qu'a prévu le législateur à plusieurs reprises comme récemment en permettant l'usage en Open Data des déclarations d'intérêts des élus à l'occasion des discussions des lois sur la transparence de la vie publique. La CNIL, en prenant des décisions qui ignorent la loi CADA (notamment sur les dossiers liés à la santé comme Sunshine) ne respecte donc pas cette hiérarchie des normes.

Ce qui oppose les démocraties des dictatures dans ce domaine est qu'en démocratie, les prises de décisions publiques sont transparentes (donc l'identité de ceux qui y participent est rendue publique) et la vie privée préservée. En dictature, la vie privée est transparente alors que les décisions publiques sont protégées.

À la lecture de certains avis de la CNIL, et au vu des difficultés que nous avons rencontrées pour les obtenir (nous avons régulièrement recours à la CADA pour obtenir des avis de la CNIL), nous avons l'impression que cette institution a tendance à privilégier l'opacité de la prise de décision publique et la protection d'acteurs puissants (par exemple les laboratoires pharmaceutiques) sur la protection des données personnelles de tous les citoyens.

### **De quelle manière la politique publique conduite par l'État, les administrations et les collectivités territoriales prend-elle en compte l'ensemble de ces exigences ?**

Les données personnelles sont un des arguments majeurs avancés en prétexte pour bloquer les processus de transparences démocratiques à l'heure de l'Internet. Si les technologies actuelles imposent sans nul doute d'assurer à tous les citoyens le respect le plus sacré de leurs informations personnelles, elles permettent également de partager largement les innombrables informations censées publiques constituées par l'administration. Il faut donc cesser d'alimenter la confusion entre Open Data et données personnelles. Si le niveau d'anonymisation n'est pas garanti, il doit être renforcé afin qu'aucune donnée personnelle ne soit divulguée à des tiers à l'exception de l'intéressé.

La réutilisation de données publiques doit être, en revanche, garantie à tous les citoyens et non réservée à une élite, que ce soit à titre gratuit ou onéreux.

### **Quels sont les dispositifs techniques ou juridiques utilisés pour garantir la protection ou la non-divulgaration, à travers l'open data, de données personnelles ?**

Le risque de divulgation de données à caractère personnel n'est pas lié à l'Open Data, il est lié à la diffusion de données non-anonymisées ou pseudo-anonymisées, dont la diffusion est généralement discrétionnaire.

Les bonnes pratiques en matière d'anonymisation sont :

- d'un point de vue juridique, le respect du secret statistique ;
- d'un point de vue technique : le recours à de la cryptographie (sha1) associée avec l'injection de chaînes aléatoires.

### **Ceux-ci présentent-ils des lacunes ? Pour quelles raisons ? Donner des exemples de failles éventuelles.**

Par essence l'Open Data exclut les données personnelles, il n'y a donc pas de tel risque dans ce cadre. La confusion entretenue entre Open Data et données personnelles et le manque d'ouverture de certaines administrations permet en revanche à certains acteurs privés d'avoir des accès privilégiés à des données personnelles de citoyens qui ne devraient pas sortir de l'administration.

### **Décrire les techniques qui permettent d'exploiter les données publiques anonymisées, pour les ré-identifier ou retrouver à partir d'elles certaines des données personnelles dont elles sont issues. Donner des exemples précis.**

Les données publiques libérées en Open Data ne permettent pas la désanonymisation en tant que telles. Seules les données pseudo-anonymisées ou mal anonymisées (donc ne correspondant pas à la définition de l'Open Data) et recoupées avec des données personnelles le permettraient.

L'accès à de nombreuses données personnelles notamment géolocalisées permet de reconnaître des individualités dans des agrégats de personnes ne respectant pas les principes du secret statistique. L'exploitation de données personnelles permet également de faciliter les techniques dites "brute force" sur les données personnelles pseudo-

anonymisées.

C'est donc le manque de contrôle autour des données personnelles qui pose problème et non l'Open Data. Sans possession de données à caractère personnel, remonter aux données personnelles de données pseudo-anonymisées est très complexe.

**Décrire les solutions qui peuvent être apportées pour contrer ces possibilités de ré-identification.**

Au vu du rôle joué par l'exploitation des données personnelles dans la ré-identification de celles pseudo-anonymisées, les bases de données à caractère personnel devraient être contrôlées de manière plus effective qu'elles ne le sont actuellement. La loi donne pourtant beaucoup de pouvoir à l'institution en charge, la CNIL. Une meilleure publicité devrait également être faite de la définition de l'Open Data, du bon respect du secret statistique et du bon usage de la cryptographie.

**Décrire les dispositifs éventuels de contrôle en matière de mise à disposition de données publiques (audit, procédure de vérification de l'impossibilité d'une ré-identification, contrôle CNIL etc.). Ceux-ci vous semblent-ils performants ?**

La commission du secret statistique est peu connue des administrations. Une des manières de rendre cette institution plus visible serait de permettre à plus d'acteurs de la solliciter pour avis, voire de prévoir un mécanisme de saisine.

La CNIL étant déjà surchargée et ayant du mal à faire respecter le droit sur les données personnelles, elle ne devrait pas être impliquée autrement que comme aujourd'hui sur les contrôles des traitement de données à caractère personnel. Elle risquerait autrement d'affaiblir encore un peu plus sa mission première : la bonne gestion des bases de données personnelles.

De plus, lors de la confection de bases de données à caractère personnel, les administrations devraient avoir l'obligation de prévoir que des exports respectant le secret statistique soient possibles.

**Comment doivent selon vous se concilier l'intérêt général qui s'attache à la diffusion des données publiques et le principe constitutionnel de protection de la vie privée ?**

Ne pas réaliser d'amalgames entre ces deux notions et s'assurer qu'elle soient clairement séparées.

**Quelles évolutions des législations CADA ou CNIL sont souhaitables de ce point de vue ?**

Il est important d'assurer la claire séparation entre l'accès à l'information publique d'un côté et le respect du droit sur les données personnelles de l'autre : c'est un vecteur de confiance citoyenne qui a démontré son efficacité. Si des institutions devaient être renforcées, il s'agirait de la commission du secret statistique afin qu'elle puisse être saisie plus largement et de la CADA afin qu'elle puisse contre-balancer les avis de la CNIL qui a tendance à privilégier les accès discrétionnaires aux données personnelles au détriment de la publicité d'informations respectant le secret statistique.

**Des améliorations de la pratique suivie en matière d'open data sont-elles nécessaires ? Lesquelles et pourquoi ?**

L'OpenData a permis de corriger des diffusions de données qui ne respectaient pas la législation en termes de protection de la vie privée et de respect du secret statistique. Il faut donc continuer dans cette voie.