Données personnelles

1. Définition

La notion de données à caractère personnel a été définie de façon extrêmement précise par le législateur en 1978, et c'est la pierre angulaire sur laquelle vient reposer toute une partie de l'arsenal législatif en matière de protection de la vie privée.

> *« Constitue une donnée à caractère personnel toute information relative à une personne physique identifiée ou qui peut être identifiée, directement ou indirectement, par référence à un numéro d'identification ou à un ou plusieurs éléments qui lui sont propres. Pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens en vue de permettre son identification dont dispose ou auxquels peut avoir accès le responsable du traitement ou toute autre personne.*

Il y a nécessairement un lien entre la donnée et la personne physique, mais ce lien peut être plus ou moins direct.

La nature des informations n'a aucune importance. Par exemple, la température de mon logement. La question clé est de savoir s'il y a un lien avec une personne physique. Si c'est le cas, la loi nous dit qu'il s'agit de données personnelles, quelle que soit la nature des données personnelles. Mais, la loi va plus loin. Ce lien peut être direct, c'est - à - dire que la base peut comporter et associer à la température mon nom. Mais, ce lien peut aussi être indirect. C'est - à - dire s'il y a un identifiant de client chez un fournisseur d'électricité par exemple (c'est ce qu'il y a derrière la notion de personne physique identifiée ou identifiable).

La loi ne se limite pas au cas où le responsable de traitement est lui - même en mesure d'identifier la personne physique. La loi parle aussi de moyens techniques auxquels aurait accès soit le responsable de traitements - c'est le cas simple - soit, toute autre personne, sous - entendue dans le monde entier. Donc, très concrètement, si EDF collecte à la fois la température d'un logement et l'identifiant de client associé, il s'agit de données personnelles. Dans ce cas, c'est assez intuitif. Mais si EDF collecte la température d'un logement et l'adresse IP du capteur de températ ure, c'est - à - dire un élément technique qui identifie un ordinateur sur Internet alors, il s'agit également de données personnelles. En effet, dans ce cas là, il existe nécessairement un fournisseur d'accès Internet qui est capable d e faire le lien entre adresse IP et client[1].

responsable de traitements : sociétés privées ou administrations publiques. L'aspect clé est que ce responsable de traitements détient une base de données qui comporte des données personnelles.

1 https://fr.wikipedia.org/wiki/Linky

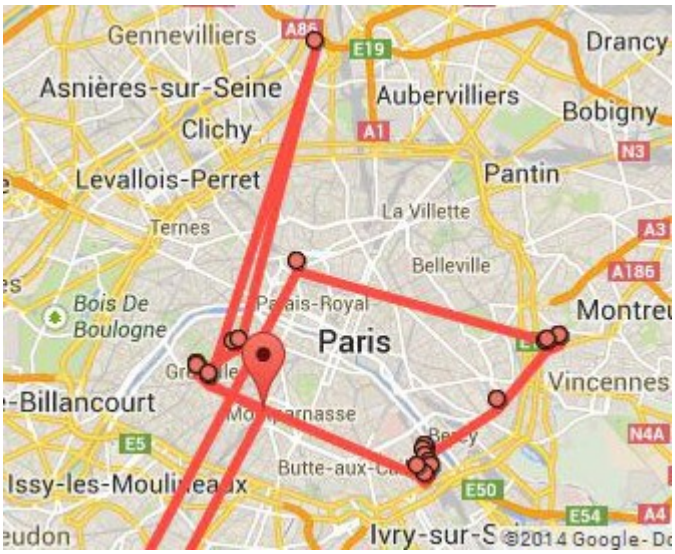
2. Données sensibles

Dans la Loi informatique et libertés, article 8, est définie une catégorie particulière de données p ersonnelles. Ce sont les données sensibles.

> *« Art. 8. - I. - Il est interdit de collecter ou de traiter des données à caractère personnel qui font apparaître, directement ou indirectement, les origines raciales ou ethniques, les opinions politiques, philosophiques ou religieuses ou l'appartenance syndicale des personnes, ou qui sont relatives à la santé ou à la vie sexuelle de celles-ci.*

Hormis certains cas particuliers, par exemple, les professionnels de la santé qui ont nécessairement besoin de manipuler des données sensibles, de façon générale, un responsable de traitements a l'interdiction totale de collecter ou d'exploiter des données sensibles.

Une difficulté de mise en œuvre est liée aux inférences. Si par exemple, un responsable de traitements, par le biais d'une application smartphone, collecte ma géolocalisation de façon systématique et précise, alors, il est capable très facilement d'inférer, c'est - à - dire de déduire un certain nombre de choses, par exemple, si je fréquente régulièrement un lieu de culte, et on tombe alors dans la catégorie des données sensibles[2].



La définitio n française qui rejoint totalement la définition européenne autour des données personnelles mais au niveau mondial le même point de vue dépend fortement de la sensibilité du pays.

Dans certains pays, le législateur, lorsqu'il s'agit de définir les données personnelles, se limite aux seuls moyens mis en œuvre par le responsable de traitements. Ainsi, le fait de collecter des données de température associées à une adresse IP ne tombe pas sous le coup de la législation autour des données personnelles, sauf si la personne, le responsable de traitements qui fait ça, la société est un fournisseur d'accès Internet. EDF, par exemple, n'a pas les moyens techniques de réidentifier une personne physique.

3. Les obligations

La loi informatique et libertés encadre très précisément les possibilités du responsable de traitements.

Certaines obligations s'imposent au responsable de traitements, dès lors qu'il manipule des données personnelles comme le stipule l'article 6 :

> *« Art. 6. - Un traitement ne peut porter que sur des données à caractère personnel qui satisfont aux conditions suivantes :*
> *« 1° Les données sont collectées et traitées de manière loyale et licite ;*

2 http://www.lemonde.fr/pixels/article/2015/07/22/google-maps-retrace-tous-vos-deplacements-dans-une-nouvelle-fonctionnalite_4694030_4408996.html

> *« 2° Elles sont collectées pour des finalités déterminées, explicites et légitimes et ne sont pas traitées ultérieurement de manière incompatible avec ces finalités. Toutefois, un traitement ultérieur de données à des fins statistiques ou à des fins de recherche scientifique ou historique est considéré comme compatible avec les finalités initiales de la collecte des données, s'il est réalisé dans le respect des principes et des procédures prévus au présent chapitre, au chapitre IV et à la section 1 du chapitre V ainsi qu'aux chapitres IX et X et s'il n'est pas utilisé pour prendre des décisions à l'égard des personnes concernées ;*
> *« 3° Elles sont adéquates, pertinentes et non excessives au regard des finalités pour lesquelles elles sont collectées et de leurs traitements ultérieurs ;*
> *« 4° Elles sont exactes, complètes et, si nécessaire, mises à jour ; les mesures appropriées doivent être prises pour que les données inexactes ou incomplètes au regard des finalités pour lesquelles elles sont collectées ou traitées soient effacées ou rectifiées ;*
> *« 5° Elles sont conservées sous une forme permettant l'identification des personnes concernées pendant une durée qui n'excède pas la durée nécessaire aux finalités pour lesquelles elles sont collectées et traitées.*

L’article 6 va imposer d'effectuer une collecte et un traitement de façon loyale et licite, mais aussi, pour des finalités bien précises et légitimes. Le responsable de traitement ne pourra pas par la suite réexploiter ces données à des fins totalement différentes. La collecte ne doit pas non plus avoir de caractère excessif au regard de la finalité. Ainsi, il est totalement hors de question de collecter ma géolocalisation en permanence si l'objectif annoncé pour cette collecte est de personnaliser un service utilisé de façon ponctuelle. Enfin, la durée de rétention de ces données personnelles ne doit pas excéder la durée nécessaire aux finalités[3].

4. La transmission de données

La transmission des données personnelles de citoyens français, européens ne peut pas être transmises hors du territoire européen. Une société américaine voulant transmettre sur des serveurs implantés sur le territoire américain des données de citoyens européens devra au préalable adhérer à un programme spécifique appelé le Privacy Shield. Dans ce programme, cette société s'engage de façon contractuelle à respecter un certain nombre de bonnes pratiques. Il s'agit d'une autocertification dans le sens où le législateur européen aura du mal à contrôler les pratiques effectives de la société. Le précédent programme qui était relativement similaire, a été invalidé par la Cour de Justice de l'Union européenne en 2014, suite aux révélations sur les pratiques de surveillance généralisée, où il apparaissait clairement que les sociétés de droit américain n'étaient pas en mesure de faire face à leurs obligations[4].

3 http://www.europe1.fr/emissions/la-une-de-leco/exploitation-de-donnees-personnelles-facebook-condamne-a-une-amende-de-110-millions-deuros-3335078
4 https://fr.wikipedia.org/wiki/Edward_Snowden

Le responsable de traitements peut échapper à ses obligations s'il obtient un consentement qualifié de libre et éclairé de la part de l'utilisateur. Si nous voyons de plus en plus apparaître des messages qui nous suggèrent de lire des règles de confidentialité, ce n'est pas pour rien. L'objectif est clairement d'essayer d'obtenir ce consentement.

Une deuxième approche pour que le responsable de traitement échappe à ses obligations est tout d'anonymiser les bases de données. Si après anonymisation, on ne peut plus faire de lien entre une donnée et une personne physique, quels que soient les moyens techniques mis en œuvre, il ne s'agit plus de données personnelles.

Dès lors, l e responsable de traitements est libre d'exploiter ces données comme bon lui semble. La difficulté est cependant d'anonymiser de façon robuste la base, et il y a eu ces dernières années beaucoup d'exemples qui ont montré que l'on pouvait parfois dé - anonymiser des bases, soi - disant anonymisées.

Un exemple de raté : publication des données d'usage d'un service de partage de vélos par les abonnés londoniens.

Il ne suffit pas de retirer le nom et de le remplacer par un identifiant aléatoire pour anonymiser une base de données... Voir (article en anglais) :

https://www.citylab.com/transportation/2014/04/londons-bike-share-program-unwittingly-revealed-its-cyclists-movements-world-see/8892/

En résumé, en ciblant un certain identifiant et en analysant son usage du service (heure et lieu de retrait et de dépôt d'un vélo), on identifie facilement la zone géographique de son lieu de résidence, de son lieu de travail, et ses habitudes. Dès lors, aller voir sur ces bornes d'emprunt aux bonnes heures permettra facilement de l'identifier avec une très forte probabilité et d'apprendre beaucoup de choses sur sa vie...

C'est une erreur grossière et il est désormais connu qu'utiliser des pseudonymes au lieu des identités véritables est à proscrire.

4. Études de cas

4.1. La société CKC

La société de carrosserie française CKC désire stocker des données clients à des fins de statistiques. Sont enregistrées "type et couleur du véhicule" ainsi que le "numéro de plaque d'immatriculation". La société fait appel à un fournisseur de stockage "Cloud" américain, qui a la totalité de ses serveurs sur le territoire américain. "Ce sont les moins chers, donc on ne va pas s'arrêter à ces petits détails" se sont-ils dit...

Elle a besoin de conseils pour être sûre de respecter le cadre légal autour des données personnelles.

1. Une voiture, c'est un tas de tôle. Ce ne sont pas des données personnelles ! Qu'en pensez-vous ?

❑ Vrai

☑ Faux

2. Comme l'hébergeur Cloud est de droit américain, c'est le droit américain qui s'applique pour les données qui se trouvent dans le Cloud ! Qu'en pensez-vous ?

   ❑ Vrai

   ☑ Faux

3. Une cliente a demandé, après son passage en caisse, combien de temps la société CKC allait conserver ses données. L'employé, derrière sa caisse, lui a répondu : "Vous savez, Madame, je ne suis qu'un employé et je ne sais pas trop... Mais ne vous en faites pas.". Est-ce correct de sa part ?

   ❑ Oui

   ☑ Non

4.2. Données anonymes à l'hôpital

Afin de permettre des études statistiques sur la fréquentation de leurs services respectifs, deux hôpitaux A et B de la région de Grenoble ont mis à disposition du public les données totalement anonymes ci dessous, indiquant pour chaque patient : le début du code postal de son lieu d'habitation, sa tranche d'âge et le service consulté.

Hôpital A			**Hôpital B**		
Code postal	**Age**	**Service**	**Code postal**	**Age**	**Service**
384**	<=30	maladies infectieuses	384**	<=35	pneumologie
384**	<=30	maternité	384**	<=35	allergologie
384**	<=30	neurologie	384**	<=35	cardiologie
384**	<=30	allergologie	384**	<=35	urologie
384**	>30	infection	384**	>35	maladies infectieuses
384**	>30	maternité	384**	>35	maternité
384**	>30	urologie	384**	>35	urologie
384**	>30	pneumologie			

L'employeur de Bob sait qu'il a 26 ans, qu'il habite dans la région de Grenoble (code postal de son lieu d'habitation: 38400), et qu'il a fréquenté les hôpitaux A puis B dans le cadre du même suivi médical. Bien sûr, il ne sait pas de quoi il s'agit. Par contre il a accès à ces deux bases de données qui sont publiques.

1. Quel est le service consulté par Bob qui peut être déduit de ces 2 bases de données ?

   ☑ allergologie

   ❑ cardiologie

- ❑ maladies infectieuses
- ❑ maternité
- ❑ neurologie
- ❑ pneumologie
- ❑ urologie

C'est un exemple trivial, mais il met en évidence que l'usage d'informations annexes (ici age, code postal et fréquentation des deux hôpitaux pour la même pathologie) permet souvent de désanonymiser des bases de données pourtant anonymes. L'anonymisation est un problème complexe, qui nécessite de trouver un compromis entre utilité et respect de la vie privée : plus les informations sont précises, plus les risques sont importants.