Fiche 03

RETOUR SOMMAIRE

Est-ce que **j'utilise des données** à caractère personnel ?

Sources : CNIL; Université Paris Nanterre

Une donnée à caractère personnel correspond à toute information se rapportant à une personne physique identifiée ou identifiable, c'est-à-dire une personne physique qui peut être identifiée directement ou indirectement.

Cette identification se fait soit par :

- → des données directement identifiantes ;
- → des données indirectement identifiantes ;
- → toute combinaison d'informations permettant d'identifier la personne.

⚠ **Une donnée à caractère personnel n'est pas forcément confidentielle ou relative à la vie privée, c'est une donnée qui a un pouvoir d'identification.**

Ex : le nom, le prénom, l'adresse IP, le son de la voix, l'image, la géolocalisation, l'historique d'un navigateur, etc.

⚠ **Une donnée n'ayant a priori pas de pouvoir d'identification, peut devenir une donnée à caractère personnel si en la croisant à une autre donnée elle permet l'identification d'une personne.**

Ex : un jeu de données comportant le lieu de résidence et la profession alors que la personne est la seule à exercer cette profession dans la localité.

Un **traitement de données à caractère personnel** est une opération (ou ensemble d'opérations) portant sur des données à caractère personnel quel que soit le procédé utilisé : collecte, enregistrement, organisation, conservation, adaptation, modification, extraction, consultation, utilisation, communication par transmission, diffusion ou toute autre forme de mise à disposition, rapprochement.

Quelques exemples de données identifiantes :

- → tous les identifiants d'une personne : nom, prénom, adresse physique ou IP, adresse mail, numéro de téléphone, numéro de sécurité sociale, numéro de compte en banque, etc. ;
- → les caractéristiques physiques : taille, poids, couleur des yeux et des cheveux, état de santé, ADN, empreintes digitales ou rétiniennes, image, son de la voix, etc. ;
- → les opinions et comportements : idées politiques, convictions religieuses, appartenances associatives, orientation sexuelle, habitudes de consommation, goûts, etc. ;
- → les données d'usage, type géolocalisation, l'image, historique de navigation ou d'achats, contenus postés, etc.

⚠ **aux spécificités des données sensibles (cf. fiche 5 : données sensibles au sens du RGPD ↗ CLIQUER ICI)**

Anonymisation / pseudonymisation

FOCUS **Le RGPD ne s'applique qu'aux données susceptibles d'identifier des personnes physiques. Si les données sont anonymisées, le RGPD ne s'applique pas. Le choix d'opérer ou non une anonymisation implique donc une réflexion rigoureuse selon le type de recherche et la sensibilité des données concernées.**

UNIVERSITÉ Clermont Auvergne

Fiche 03

Est-ce que j'utilise des données à caractère personnel ?

Sources : CNIL, Université Paris Nanterre

Définitions CNIL

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre impossible, en pratique, toute identification de la personne par quelque moyen que ce soit et ce de manière irréversible.

La pseudonymisation est un traitement de données à caractère personnel réalisé de manière à ce qu'on ne puisse plus attribuer les données relatives à une personne physique sans avoir recours à des informations supplémentaires.

Contrairement à l'anonymisation, elle permet une ré-identification des personnes.

- En pratique la pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro dans un classement, etc.). La pseudonymisation permet ainsi de traiter les données d'individus sans pouvoir identifier ceux-ci de façon directe. Il est toutefois possible de retrouver l'identité de ceux-ci grâce à des données tierces. C'est pourquoi des données pseudonymisées demeurent des données à caractère personnel.

Comment choisir les techniques d'anonymisation ?

Le processus d'anonymisation vise à éliminer toute possibilité de ré-identification : il implique donc une nécessaire perte de qualité des données. Leur exploitation future est ainsi limitée à certains types d'utilisation. Ces contraintes sont à prendre en compte dès le début du projet.

Pour construire un processus d'anonymisation pertinent, il est ainsi conseillé par la CNIL de :

- supprimer les éléments d'identification directe ainsi que les valeurs rares qui pourraient permettre une ré-identification aisée des personnes (par exemple, la connaissance précise de l'âge des individus) ;
- distinguer les informations importantes des informations secondaires ou inutiles (c'est-à-dire supprimables) ;
- définir le degré d'anonymat idéal et acceptable pour chaque information conservée ;
- définir les priorités (par exemple, est-il plus important de conserver une grande finesse sur telle information ou de conserver telle autre information ?).

Ce questionnement aide à déterminer le procédé d'anonymisation le plus pertinent :

- **la randomisation** consiste à modifier les attributs dans un jeu de données de telle sorte qu'elles soient moins précises, tout en conservant la répartition globale. Cette technique permet de protéger le jeu de données du risque d'inférence (exemple : permuter les données relatives à la date de naissance des individus de manière à altérer la véracité des informations contenues dans une base de données) ;
- **La généralisation** permet de généraliser les attributs du jeu de données en modifiant leur échelle ou leur ordre de grandeur afin de s'assurer qu'ils soient communs à un ensemble de personnes. Cette technique permet d'éviter l'individualisation d'un jeu de données. Elle limite également les possibles corrélations du jeu de données avec d'autres (exemple : dans un fichier contenant la date de naissance des personnes, il est possible de remplacer cette information par la seule année de naissance, ou une fourchette).

Chaque technique d'anonymisation présente ses avantages et sera à décider en fonction du traitement de données et de l'objectif poursuivi.

Trois critères cumulatifs permettent d'évaluer l'efficacité d'une solution d'anonymisation :

- **l'individualisation :** est-il toujours possible d'isoler un individu ?
- **la corrélation :** est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?
- **l'inférence :** peut-on déduire de l'information sur un individu ?

un ensemble de données pour lequel il n'est possible ni d'individualiser ni de corréler ni d'inférer est *a priori* anonyme.

FOCUS

La Commission européenne dans le cadre du programme H2020 propose aux chercheurs l'outil « Amnesia », une solution applicable à des données recueillies à l'état brut permettant une anonymisation des résultats.

https://amnesia.openaire.eu/