LA DIFFICILE ANONYMISATION DES DONNEES PERSONNELLES

Chaire Valeurs et Politiques
DES INFORMATIONS PERSONNELLES

Maryline Laurent, professeur à Télécom SudParis et co-fondatrice de la chaire Valeurs et politiques des informations personnelles, et Claire Levallois-Barth, Coordinatrice de la Chaire Valeurs et politiques des informations personnelles, Maître de Conférences en droit à Télécom Paristech

Une des solutions techniques souvent mise en avant pour répondre à la problématique de protection du citoyen est celle de l'anonymisation des données personnelles. Cette solution est aujourd'hui essentielle pour de nombreux acteurs qui souhaitent valoriser les informations qu'ils détiennent, notamment dans le cadre de l'open data ou de l'Internet des objets. Pour autant, il est souvent difficile de savoir comment anonymiser convenablement des données.

A cet égard, le groupe de l'Article 29 (G29) qui regroupe les autorités de protection des données personnelles de l'Union européenne, dont la CNIL, a publié le 10 avril 2014 un avis sur les principales techniques d'anonymisation [G29]. Le document distingue la **donnée anonymisée** et la **donnée pseudonymisée.** Sur le plan juridique, la donnée pseudonymisée n'est pas une donnée anonyme dans la mesure où elle ne constitue pas une dé-identification irréversible et permet de remonter à la personne concernée ; la donnée pseudonymisée reste donc une donnée personnelle.

L'avis explique également comment mettre en œuvre une solution d'anonymisation. Pour l'essentiel, les techniques d'anonymisation font appel à deux grand principes :

- **La randomization** consiste à altérer les données personnelles par exemple par des techniques de bruitage pour distendre le lien entre les données et la personne réelle.
- **La généralisation** consiste à généraliser ou diluer les données personnelles de façon à ce qu'elles perdent en précision et qu'elles ne soient plus spécifiques à une personne mais communes à un ensemble de personnes.

Chaque technique présente des avantages et inconvénients que le G29 propose d'évaluer à travers trois critères : la ré-identification, la corrélation et l'inférence. Une technique permet rarement à elle seule d'anonymiser réellement un jeu de données. Un procédé effectif d'anonymisation doit par conséquent définir la combinaison de techniques qui répond le mieux à l'objectif recherché. En règle générale, ce procédé est basé sur au moins deux techniques d'anonymisation. Adapté au cas par cas, il doit tenir compte de l'existence des différents jeux de données disponibles par ailleurs et des possibilités de croisement entre ces jeux de données et le jeu que l'on souhaite anonymiser.

Dans un contexte de Big Data, l'anonymisation est d'autant plus difficile à garantir que le volume de données disponibles évolue très vite et de façon massive. Or, le procédé d'anonymisation exige du temps pour sa mise au point et l'évaluation de son niveau de fiabilité. Ce travail d'évaluation peut s'avérer difficile du fait du gros volume de données disponibles, et de la nécessité de s'assurer que l'introduction d'un nouveau jeu de données ne permettra pas de ré-identifier une personne. En pratique, le respect des règles juridiques est susceptible d'aboutir à l'introduction dans le Big Data de jeux de données anonymisés de moins en moins précis et pertinents.

Sur le plan juridique, le procédé d'anonymisation constitue en lui-même un traitement de données personnelles. Il doit par conséquent être conforme aux obligations posées par la directive européenne Protection des données personnelles [DIR 95], transposée en droit français par la loi Informatique et Libertés [LOI 78-17]. Cette législation devrait dans un futur proche (certainement avant fin 2015) être remplacée par un règlement [PROP REG 2012]. D'application directe, le nouveau texte ne nécessitera pas de transposition en droit national : la loi Informatique et Libertés sera abrogée tandis que la CNIL interprétera directement le nouveau règlement. L'avis du G29 (qui prendra le nom de Comité européen de la protection des données) sur les techniques d'anonymisation restera pertinent et ce, d'autant plus que la dernière version du projet de règlement distingue bien les données anonymisées des données pseudonymisées [PROP REG 2015].

Les lecteurs intéressés pourront se référer au glossaire consacré aux notions d'anonymat/pseudonymat défini dans le cadre de la Chaire de l'Institut Mines-Télécom Valeurs et politiques des informations personnelles, dont un extrait est fourni ci-dessous :

Anonymisation.- Processus par lequel des informations personnellement identifiables sont altérées de façon irréversible de sorte que la personne à laquelle se rapporte l'information ne peut plus être identifiée directement ou indirectement [ISO 29100].

Données à caractère personnel.- Toute information concernant une personne physique identifiée ou identifiable (personne concernée); est réputée identifiable une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques, propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale (Art. 2 a de la directive Protection des données [DIR 95]).

Donnée anonymisée.- Donnée qui est le résultat d'un processus d'anonymisation d'informations personnellement identifiables [ISO 29100].

Donnée pseudonymisée.- Donnée qui a subi un traitement de pseudonymisation.

Généralisation.- Technique d'anonymisation qui consiste à généraliser ou diluer les données personnelles par modification de leur échelle ou de leur ordre de grandeur. Il s'agit par exemple de remplacer la date de naissance par l'année de naissance, la ville par la région [G29].

Pseudonymisation.- Traitement appliqués sur des données à caractère personnel de manière que ces données ne puissent pas être associées à la personne concernée sans l'utilisation d'informations supplémentaires, et ce tant que ces informations supplémentaires sont conservées séparément et soumises à des mesures techniques et organisationnelles garantissant la non-attribution des données à une personne identifiée ou identifiable (Article 4 (3b) [PROP REG 2012]). **Randomization**. Ensemble de techniques d'anonymisation qui altèrent la véracité des données dans le but de supprimer le lien fort entre les données et la personne [G29].

Ré-identification.- Action consistant à isoler plusieurs voire tous les enregistrements qui identifient un individu dans un ensemble de données [G29].

[DIR 95] Directive 95/46/CE du Parlement européen et du Conseil, du 24 octobre 1995, relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, JOCE n° L 281 du 23 novembre 1995

[G29] ARTICLE 29 DATA PROTECTION WORKING PARTY, Opinion 05/2014 on Anonymisation Techniques adopted on 10 April 2014, WP216 disponible uniquement en anglais.

[ISO 29100] International Standard, *Information technology - Security techniques - Privacy framework*, ISO/IEC29100, First edition, Décembre 2011.

[LOI 78-17] Loi n° 78-17 du 6.01.1978 relative à l'informatique, aux fichiers et aux libertés, JORF 7.01.1978, p. 227 (Loi dite Informatique et Libertés).

[PROP REG 2012] Proposition de règlement du Parlement européen et du Conseil relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, COM(2012)11 final

[PROP REG 2015] Proposition de règlement du Parlement européen et du Conseil relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, Council's consolidated version of March 2015