L’anonymisation, une fausse bonne idée ?

Nathanaël Pascal, le 13 novembre 2024

La CNIL observe et retient le caractère non anonyme des données de santé traitées par une société spécialisée dans l’édition et la vente de logiciels de gestion à destination de professionnels de la santé. Il en résulte des manquements à l’obligation d’effectuer les formalités préalables dans le domaine de la santé, à savoir de formuler une demande d’autorisation auprès de la CNIL ou lui adresser une déclaration de conformité à l’un de ses référentiels, ainsi qu’à l’obligation de traiter les données de manière licite, lesquels sont lourdement sanctionnés.

Délibération de la formation restreinte n°SAN-2024–013 du 5 septembre 2024 concernant la société CEGEDIM SANTÉ

En pratique, il est fréquent que des responsables de traitement se targuent de procéder à l’anonymisation des données afin de se soustraire à l’application de la réglementation sur la protection des données. En effet, l’anonymisation ouvre des potentiels de réutilisation initialement prescrits du fait du caractère personnel des informations. Toutefois, nombreux sont les cas dans lesquels l’anonymisation dont se prévaut le responsable du traitement est uniquement constitutive de pseudonymisation avec pour conséquence l’application de la réglementation et ses nombreuses incidences.

À titre d’exemple, l’autorité de protection des données française (CNIL) a sanctionné une société spécialisée dans l’édition et la vente de logiciels de gestions à destination des professionnels de la santé (ci-après : Société). Cette Société offre à ses clients la possibilité d’accéder à une base de données qui regroupe des informations relatives aux patients, et ce, dans le but de leur permettre de conduire des études et des analyses statistiques dans le domaine de la santé. Les données figurant dans la base de données concernée étaient extraites des dossiers médicaux des patients suivis par les médecins recourant au logiciel développé par la Société. Ces derniers avaient consenti à la transmission des données dans le cadre d’un observatoire épidémiologique institué par la société éditrice. En contrepartie, les professionnels de la santé bénéficiaient d’une réduction sur la licence d’utilisation dudit logiciel, qui leur permet d’accéder aux études statistiques réalisées par les autres clients de la Société. En 2021, la CNIL a conduit des enquêtes au cours desquelles elle a découvert que, dans le cadre de l’utilisation dudit logiciel, la Société avait traité sans autorisation des données qui

n'étaient pas anonymes, la réidentification des personnes concernées étant techniquement possible.

La formation restreinte de la CNIL traite préalablement des notions de pseudonymisation et d'anonymisation afin de formuler à l'égard de la Société deux manquements à la réglementation applicable à savoir la violation de l'obligation d'effectuer les formalités préalables dans le domaine de la santé et la violation du principe de licéité. La formation restreinte conclut au prononcé d'une amende administrative de EUR 800'000.- pour des faits s'étant achevés en 2022.

Du caractère non anonyme des données traitées

Dans le cadre de son observatoire, la Société collectait des données relatives au dossier administratif du patient, au dossier médical, aux prescriptions pharmaceutiques et aux autres prescriptions. Chaque patient se voyait attribué un identifiant unique dans le flux du logiciel édité par la Société et dans les fichiers transmis à cette dernière. En effet, toutes les données d'un même patient suivi par un même médecin étaient associées au même identifiant dans la base de données. Le recours à cet identifiant unique permet, selon la Société, de ne pas être soumise au régime applicable en matière de protection des données puisque les données traitées sont anonymes.

La formation restreinte de la CNIL rappelle les notions de données à caractère personnel (art. 4 par. 1 RGPD) et de santé (art. 4 par. 15 RGPD) et constate que, contrairement à celle de pseudonymisation (art. 4 par. 5 RGPD), la notion d'anonymisation n'est pas définie par le RGPD. Dès lors, la formation restreinte s'intéresse à la jurisprudence rendue par la CJUE. Elle cite l'arrêt Breyer (C-582/14, § 42–43) dont l'apport consiste principalement dans le fait que pour déterminer si une personne est identifiable, il convient de considérer l'ensemble des moyens susceptibles d'être raisonnablement mis en œuvre pour l'identifier, étant précisé qu'il n'est pas requis que toutes les informations permettant une telle identification se trouvent entre les mains d'une seule personne.

Dans un autre arrêt (C-479/22, commenté *in* swissprivacy.law/304/), la CJUE a précisé que le fait que des informations supplémentaires soient nécessaires à l'identification n'exclut pas que les données en question puissent être qualifiées de données personnelles (arrêt précité, § 49). La possibilité de combiner ces données avec des informations supplémentaires constitue un moyen raisonnablement envisageable pour identifier la personne concernée. Pour ce faire, il convient de tenir compte de l'ensemble des facteurs objectifs tels que le coût de l'identification et le temps nécessaires à celle-ci, ainsi que les technologies disponibles au

moment du traitement et l'évolution de celles-ci (arrêt précité, § 50). Au surplus, la Cour précise qu'il est inhérent à l'identification indirecte d'une personne que des informations supplémentaires soient combinées avec les données en question afin d'identifier la personne concernée (arrêt précité, § 55). Dite identification dont la preuve n'a pas à être apportée en raison du niveau d'exigence règlementaire se limitant à exiger qu'elle soit identifiable (arrêt précité, § 61). Quand bien même cette exigence figure au Règlement (UE) 2018/1725, la CNIL procède à un raisonnement par analogie en se basant sur le fait que la définition est identique à celle figurant à l'art. 4 par. 1 RGPD.

Last but not least, la CJUE (C-604/22, commenté *in* swissprivacy.law/303/) a jugé qu'une chaîne composée une combinaison de lettres et de caractères contenant les préférences d'un utilisateur d'internet constitue une donnée personnelle. La Cour est arrivée à une telle conclusion indépendamment du fait qu'à défaut de contribution extérieure, l'organisation sectorielle détenant ladite chaîne ne pouvait ni accéder aux données traitées par les membres du *framework*, ni combiner ladite chaîne avec d'autres éléments.

En l'absence de référentiels et de lignes directrices spécifiques publiés par la CNIL, sa formation restreinte s'est tournée vers un avis adopté par le Groupe de travail « article 29 » il y a plus de dix ans, lequel indique qu'un processus peut être qualifié d'anonymisation lorsqu'il parvient à résister aux risques suivants :

> *"- l'individualisation*, qui correspond à la possibilité d'isoler une partie ou la totalité des enregistrements identifiant un individu dans l'ensemble de données ;
>
> - *la corrélation*, qui consiste dans la capacité de relier entre elles, au moins, deux enregistrements se rapportant à la même personne concernée ou à un groupe de personnes concernées (soit dans la même base de données, soit dans deux bases de données différentes). Si une attaque permet d'établir (…) que deux enregistrements correspondent à un même groupe d'individus, mais ne permet pas d'isoler des individus au sein de ce groupe, la technique résiste à l'« individualisation », mais non à la corrélation ;
>
> - *l'inférence*, qui est la possibilité de déduire, avec un degré de probabilité élevé, la valeur d'un attribut à partir des valeurs d'un ensemble d'autres attributs".

La CNIL en conclut que lorsque le responsable du traitement se prévaut de l'anonymisation de données, il lui appartient de démontrer que la réidentification par le biais de moyens

raisonnables n'est pas possible, de sorte que de tels risques sont négligeables.

En l'espèce, la Société faillit à cet exercice car elle est en mesure, grâce à l'identifiant du patient, de relier ledit identifiant à divers fichiers transmis successivement par un même médecin concernant un même patient, offrant ainsi à la Société une vision du parcours de soin. Le traitement ne résistant pas au risque de l'individualisation, la formation restreinte procède à l'évaluation *in concreto* du risque de réidentification. Dans ce cadre, elle constate qu'il est possible de réidentifier un nombre important d'individus dans un jeu de données pseudonymisées à partir de données de géolocalisation. Une corrélation entre les données tierces et les informations détenues par la Société augmenterait considérablement les possibilités de levée du pseudonymat. Or, le rapporteur de la CNIL est parvenu à isoler un individu et à suivre partiellement son parcours de soin sans avoir à recourir à des informations supplémentaires de sorte qu'il est possible de lever le pseudonymat dans le cas d'espèce.

Par conséquent, la formation restreinte retient que malgré l'attribution d'un identifiant unique, la richesse des informations détenues par la Société était si importante que la levée du pseudonymat était possible. Dès lors, la CNIL tire les conséquences de l'absence d'anonymisation.

De la notion d'entrepôt de données de santé et des formalités préalables dans le domaine de la santé

À titre préliminaire, il convient de préciser que la notion « d'entrepôt de données de santé » est une émanation doctrinale de la CNIL, laquelle s'apprécie selon un faisceau d'indices. Parmi ceux-ci, nous pouvons retrouver la durée de conservation, la réutilisation dans des traitements ultérieurs des données, l'alimentation au fil de l'eau de la base de données ainsi que les finalités du traitement. La formation restreinte estime que la Société a constitué un entrepôt de données de santé en procédant à une collecte massive et au fil de l'eau de données issues des dossiers médicaux, et ce, en vue de les mettre à la disposition de tiers de manière mutualisée afin qu'ils puissent traiter ces données à des fins d'études et de recherches dans le domaine de la santé. Or, selon l'art. 66 de la loi Informatique et Libertés (dite « LIL ») les traitements de données personnelles dans ce domaine, dont font partie les entrepôts de données de santé, ne peuvent être mis en œuvre qu'après autorisation de la CNIL ou à la condition d'être conformes à l'un de ses référentiels.

Estimant à tort avoir anonymisé les données concernées, la Société n'a pas recueilli le consentement explicite des personnes concernées (art. 65 1° LIL), ne s'est pas conformée à un référentiel ou sollicité une autorisation auprès de la CNIL, traitant de la sorte des

données de santé sans autorisation.

Du principe de licéité

Lors de la conduite de ses enquêtes, la CNIL a constaté qu'une collecte automatique de données était opérée lors de l'utilisation d'un téléservice de l'assurance maladie permettant d'accéder à l'historique des remboursements de santé effectués au cours des douze derniers mois.

La formation restreinte retient que la Société a manqué à son obligation de traiter les données de manière licite (art. 5 par. 1 let. a RGPD) en n'offrait pas d'alternative, comme la consultation desdites données, à ses utilisateurs.

Conclusion

Cette décision illustre la nécessité impérieuse de garder à l'esprit que l'anonymisation est ardue à garantir. Sa reconnaissance peut notamment dépendre du nombre de données traitées relatives à une personne concernée ou encore du processus retenu. En effet, le responsable du traitement doit conserver à l'esprit que le processus d'anonymisation valable à l'instant *T* se devra d'également résister de manière anticipée à des technologies n'étant pas encore disponibles. De ce fait, le processus d'anonymisation adopté devra notamment faire l'objet d'une veille technologique afin de prévenir d'éventuelles modifications. Par conséquent, les tentatives d'anonymisation sont fréquemment requalifiées par les autorités de protection des données en pseudonymisation, notamment en raison du risque de corrélation.

Proposition de citation : Nathanaël PASCAL, L'anonymisation, une fausse bonne idée ?, 13 novembre 2024 *in* www.swissprivacy.law/323

Les articles de swissprivacy.law sont publiés sous licence creative commons CC BY 4.0.